



Pettitt, Michael Andrew (2008) Visual demand evaluation methods for in-vehicle interfaces. PhD thesis, University of Nottingham.

Access from the University of Nottingham repository:

http://eprints.nottingham.ac.uk/10436/1/Pettitt_2008_PhD_Thesis.pdf

Copyright and reuse:

The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

- Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners.
- To the extent reasonable and practicable the material made available in Nottingham ePrints has been checked for eligibility before being made available.
- Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.
- Quotations or similar reproductions must be sufficiently acknowledged.

Please see our full end user licence at:

http://eprints.nottingham.ac.uk/end_user_agreement.pdf

A note on versions:

The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact eprints@nottingham.ac.uk

**VISUAL DEMAND EVALUATION METHODS FOR IN-VEHICLE
INTERFACES**

Michael Pettitt, BA (Hons) MSc

**Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy**

February 2008

Abstract

Advancements in computing technology have been keenly felt in the automotive industry. Novel in-car systems have the potential to substantially improve the safety, efficiency and comfort of the driving experience. However, they must be carefully designed, so their use does not dangerously distract drivers from fundamental, safety-critical driving tasks. Distraction is a well-established causal factor in road accidents. A concern is that the introduction of new in-vehicle technology may increase exposure to distraction, and lead to an increase in distraction-related accidents. The range of systems often termed In-Vehicle Information Systems (IVIS), encompassing navigation and entertainment systems, in-car email and Internet, are the focus of this thesis, since they are commonly associated with long tasks that are not considered fundamentally relevant to driving.

A variety of Human-Computer Interaction (HCI) and Human Factors methods has been employed to assess the potential distraction of IVIS task engagement. These include on-road evaluations, driving simulator studies, and surrogate methods, such as peripheral detection tasks and static task time assessments. The occlusion technique is one such surrogate, where task performance is assessed under intermittent vision conditions. Participants complete a task with 1.5-second vision periods, followed by a period where their vision is occluded. In this way, the technique evaluates how visually demanding a task is, mimicking the behaviour of glancing to and from the forward road scene when driving and performing IVIS tasks. An evaluation of the technique's validity is presented. Sixteen participants performed two tasks on two systems under three conditions: static (full-vision), static (occlusion), and, whilst driving. Results confirmed other research, concluding that the technique is valid. However, the method's assessment through user-trials based on measures of human performance is problematic. Such trials require robust, reliable prototype systems, and can therefore only take place in later design stages. Consequently, the economic effectiveness of the technique is questionable.

The keystroke-level model (KLM), which predicts task times for error-free performance by expert users in routine tasks, provides an alternative quantitative assessment method to user-trials. Tasks are decomposed into their most primitive actions, termed operators, which are associated with empirically assessed time values. These values are then summed to predict performance times. An evaluation of the technique in a vehicle environment is presented; twelve participants performed eleven tasks on two in-car entertainment systems, and task times were compared with KLM predictions. Results demonstrate the technique remains valid beyond its original, desktop computing based context. However, the traditional KLM predicts static task time only, and an extended procedure is required to consider occluded task performance.

Two studies are presented, seeking to extend the KLM in order to model task performance under the interrupted vision conditions of occlusion trials. In the

first, predictions of occlusion metrics are compared with results from the earlier occlusion assessment. In the second, twelve participants performed three tasks on two IVIS systems under occlusion conditions. Results were subsequently compared with predicted values. Both studies conclude that the extended KLM approach produces valid predictions of occlusion methods, with error rates generally within 20% of observed values, in line with expectations for KLM predictions. Subsequently, a case study is presented, to demonstrate the technique's reliability. The results of an independent occlusion study of two IVIS tasks are compared with predictions made by a HCI expert trained in the application of the extended KLM. Error rates for this study were equally low, leading to the conclusion that the extended KLM appears reliable, though further studies are required.

It is concluded that the extended-KLM technique is a valid, reliable and economical method for assessing the visual demand of IVIS tasks. In contrast to many user-trial methods, the technique can be applied in early design stages. In addition, future work areas are identified, which could serve to further enhance the validity, reliability and economy of the technique. These include, automating the extended KLM procedure with a software tool, and, the development of new cognitive and physical operators, and new assumptions, specific to IVIS and/or occlusion conditions. For example, it will be useful to develop new cognitive operators that consider the time taken to visually reorient to complex displays following occluded periods.

Publications

Over the course of this research the author has published the following papers.

- PETTITT, M. A., BURNETT, G. E., BAYER, S. & STEVENS, A. (2006)** Assessment of the occlusion technique as a means for evaluating the distraction potential of driver support systems. *IEE Proceedings on Intelligent Transport Systems*, 4, 259-266.
- PETTITT, M. A., BURNETT, G. E. & KARBASSIOUN, D. (2006)** Applying the keystroke level model in a driving context. IN BUST, P. D. (Ed.) *Contemporary Ergonomics: Proceedings of the Ergonomics Society Annual Conference 2006*. Cambridge, UK, Taylor and Francis, 219-223.
- PETTITT, M. A., BURNETT, G. E. & STEVENS, A. (2005)** Defining driver distraction. *Proceedings of 12th ITS World Congress*. San Francisco, USA, ITS America.
- PETTITT, M. A., BURNETT, G. E. & STEVENS, A. (2006)** Extending the keystroke level model (KLM) to assess the visual demand of an in-vehicle information system (IVIS). *Proceedings of 13th World Congress on Intelligent Transport Systems 2006*. London, UK, Ertico.
- PETTITT, M. A., BURNETT, G. E. & STEVENS, A. (2007)** An extended keystroke level model (KLM) for predicting the visual demand of in-vehicle information systems. Full paper, *CHI'2007*. San Jose, California, USA, ACM, 1515-1524.
- PETTITT, M., BURNETT, G. & STEVENS, A. (2007)** An extended KLM for predicting the visual demand of in-vehicle information systems. IN: BUST, P. D. (Ed.) *Contemporary Ergonomics: Proceedings of the annual conference of the Ergonomics Society*. Cambridge, UK, Taylor and Francis, 272-277.

Acknowledgements

The research presented in this thesis would not be in the format it is, and in many ways, it would not have been possible at all, without the help, support and encouragement of the following people. First, I must extend my thanks to my supervisor, Dr. Gary Burnett. His knowledge, expertise, encouragement, direction and assistance have been invaluable. Secondly, thanks go to my secondary supervisor, at the Transport Research Laboratory, Dr. Alan Stevens. His fresh perspective and positive attitude has been a driving force behind the work.

A number of people have assisted in the development and running of the trials and studies described within. For their assistance in data collection, I would like to thank: Steven Bayer and Terry Harrison of ESRI, and, Alex Irune and Darius Karbassioun of the University of Nottingham. I would also like to thank John Richardson of ESRI, Mark Porter and Steve Summerskill of Loughborough University, Katharine Prynne of Honda, Tim Horberry of TRL, and Prof. David Clarke of the University of Nottingham, for providing advice on study design, and/or access to equipment or data.

Finally, I wish to thank my friends and family for their constant love, support and encouragement. I thank my colleagues at the MRL at the University of Nottingham, and my new colleagues in the Human Factors Research Group, particularly Ash Patel, Rose Saikayasit and Glyn Lawson. Thanks especially to my parents for their belief in me, and for never expecting too much! And to my friends, who have been a great support throughout: Alex Colston, Craig Danvers, Peter Hextall, Matthew Ibbotson, Rebecca Mayes, Mark Terry, and anyone else who I've forgotten to mention.

List of Contents

Abstract.....	ii
Publications.....	iv
Acknowledgements.....	v
List of Contents.....	vi
List of Tables.....	x
List of Figures.....	xi
List of Abbreviations.....	xii
Preface.....	xiv
1 Introduction.....	1
1.1 Cars and Technology.....	1
1.2 Distraction and accident rates.....	5
1.3 HCI methods and driver distraction.....	8
1.4 Aims and Objectives.....	12
1.5 Research Contribution.....	13
1.6 Structure of the thesis.....	14
2 Literature review: driver distraction definitions and accident statistics ...	20
2.1 Driver distraction: current definitions.....	20
2.1.1 Distraction and inattention.....	23
2.1.2 Internal and external distraction.....	25
2.1.3 Types of distraction.....	27
2.1.4 Distracting activities and their relationship to the primary driving task.....	28
2.2 Working definition.....	30
2.3 Distraction as a factor in accident causation.....	31
2.3.1 Case-control studies.....	33
2.3.2 Results from around the world.....	36
2.3.3 Results from the UK.....	37
2.3.4 Sources of distraction.....	39
2.3.5 The role of technology in distraction-related accidents.....	41
2.3.6 Types of distraction.....	48
2.4 Summary.....	49
3 Literature review – Measuring distraction: evaluation methods.....	51
3.1 User trial methods.....	52
3.1.1 The 15-second rule.....	54
3.1.2 The occlusion technique.....	56
3.1.3 Lane change test (LCT).....	68
3.1.4 Peripheral detection task.....	72
3.1.5 Simulator and on-road studies.....	73
3.2 Non-user trial methods.....	83
3.2.1 Qualitative techniques.....	83
3.2.2 HCI and modelling behaviour.....	84
3.2.3 GOMS and KLM.....	85
3.2.4 Using software tools to automatically generate KLMs.....	90
3.2.5 Predicting distraction through cognitive models of driver behaviour92	
3.3 Opportunities for research.....	96
3.4 Summary.....	100

4	Study 1 – Validation of the occlusion technique	102
4.1	Research Questions	103
4.2	Research contribution	104
4.3	Method	104
4.3.1	Participants	104
4.3.2	Equipment	105
4.3.3	Tasks	106
4.3.4	Trial set-up and procedure	106
4.4	Results	108
4.4.1	Results from occlusion trials and on-road evaluation.....	108
4.5	Discussion.....	112
4.5.1	Validity of technique	112
4.5.2	Ease of implementation	114
4.6	Summary.....	116
5	Study 2 – Applying the keystroke level model in a driving context.	117
5.1	Research questions.....	119
5.2	Research contribution	119
5.3	Method.....	119
5.4	Results	121
5.4.1	Correlations between observed and predicted task times	121
5.4.2	Error rates	123
5.5	Discussion.....	124
5.6	Summary.....	127
6	Study 3 – Feasibility study: extending the keystroke level model to assess the visual demand of in-vehicle information systems.	128
6.1	Research questions.....	129
6.2	Research contribution	129
6.3	Method.....	130
6.3.1	Tasks	130
6.3.2	Occlusion trials	131
6.3.3	KLM models	133
6.4	Comparison between methods.....	136
6.4.1	Error rates	138
6.5	Discussion.....	138
6.6	Summary.....	141
7	Study 4 – Validation of the extended KLM	142
7.1	Research questions.....	143
7.2	Research contribution	143
7.3	Method.....	143
7.3.1	Occlusion	143
7.3.2	Extended KLM	147
7.4	Comparison between predicted and observed results.....	154
7.5	Discussion.....	154
7.6	Summary.....	156
8	Case Study – Reliability of the Extended KLM	157
8.1	Research questions.....	159
8.2	Research contribution	159
8.3	Tasks	159
8.4	Occlusion trials	160
8.4.1	Participants	160

8.4.2	Equipment.....	160
8.4.3	Trial procedure.....	161
8.4.4	Results	161
8.5	Extended KLM predictions.....	162
8.5.1	Modeller.....	162
8.5.2	Procedure	162
8.6	Findings	166
8.7	Discussion.....	167
8.7.1	Reliability and validity	167
8.7.2	Economy	170
8.8	Summary.....	171
9	General Discussion.....	172
9.1	IVIS and driver distraction	172
9.2	Occlusion	174
9.3	Keystroke Level Model	175
9.4	Extended KLM	177
9.4.1	Validity	177
9.4.2	Reliability	180
9.4.3	Economy	182
9.5	Summary of extended KLM future work	183
9.6	Contribution to research	185
9.6.1	The extended KLM in the context of other evaluation methods 186	
9.6.2	Use of the extended KLM in other contexts.....	190
9.6.3	Implications for IVIS design	191
9.7	Summary.....	192
10	Conclusions	194
10.1	Distraction and accidents (Objective 1).....	194
10.2	Evaluation methods (Objective 2)	195
10.3	Development of extended KLM (Objective 3).....	196
10.4	Summary.....	199
	References.....	201
	Appendix A – Prevalence of driver distraction: results from the work related road traffic accident database.	211
	Work related road traffic accident database	211
	Method.....	213
	Results and discussion	213
	Distraction and Inattention	213
	Internal and external distraction	218
	Types of distraction	221
	Distraction and the driving task	222
	Distraction and accident type.....	223
	Summary.....	225
	Appendix B: Example Consent Forms	227
	Study 1	227
	Study 4	228
	Appendix C: Participant questionnaires	230
	Study 1	230
	Study 4.....	232
	Appendix D: Raw data.....	233

Study 1	233
Study 2	238
Study 4	240

List of Tables

Table 1.1 Study details	15
Table 2.1 Definitions of distraction	22
Table 2.2 Summary of weighted CDS data for driver distraction	40
Table 3.1 Driving performance variables and measures	74
Table 4.1 Participant driving experience	104
Table 4.2 Participant data numbers for each condition	108
Table 4.3 Comparison between systems for results from occlusion and on-road trials	111
Table 4.4 Comparison between tasks within systems for results from occlusion and on-road trials	111
Table 5.1 Tasks	120
Table 5.2 Results	123
Table 5.3 Prediction error rates for all tasks on both systems	123
Table 6.1 Summary of occlusion statistics from data gathered during user trials	133
Table 6.2 Selection of operators used in KLM models	134
Table 6.3 Summary of extended KLM predictions for TSOT and R	136
Table 6.4 Prediction error rates for TSOT and R values	138
Table 7.1 Occlusion results	147
Table 7.2 KLM predictions of static task time, TSOT and R	154
Table 7.3 Error rates of KLM predictions	154
Table 8.1 Mean TSOT and R observed values	161
Table 8.2 Summary of observed and predicted results	167

List of Figures

Figure 1.1 The impact of technology on the driving experience	3
Figure 1.2 Categorisation framework for evaluation methods	12
Figure 1.3 Structure of the thesis	16
Figure 3.1 Experimental set up	58
Figure 3.2 Occlusion goggles: (a) shutters open; (b) shutters closed	59
Figure 3.3 Lane-change task set-up	70
Figure 3.4 Example screen in LCT	70
Figure 3.5 Example of lane-change task performance	70
Figure 3.6 Lateral deviation results for input modality validation study	95
Figure 3.7 Lateral velocity results for age effects study	95
Figure 3.8 Evaluation methods by economy and validity	97
Figure 3.9 Evaluation methods according to validity, economy and reliability	98
Figure 4.1 Camera views	106
Figure 5.1 Comparison of mean observed task time and predicted task time for System 1	122
Figure 5.2 Comparison of mean observed task time and predicted task time for System 2	122
Figure 6.1 Example of modelling occluded performance	136
Figure 6.2 Comparison of predicted and observed TSOT values	137
Figure 6.3 Comparison of predicted and observed R-values	138
Figure 7.1 System A (left) and System B (right)	144
Figure 7.2 KLM operator timelines	153
Figure 8.1 Extended KLM for POI search task	163
Figure 8.2 Extended KLM for destination entry task	164
Figure 8.3 Extended KLM for short scrolling text task	166
Figure 8.4 Extended KLM for long scrolling text task	166
Figure 9.1 Extended KLM within the evaluation framework	187
Figure 9.2 Evaluation methods according to validity and economy	188
Figure 9.3 Driver Workload Metrics Toolkit	189

List of Abbreviations

AAM	Alliance of Automobile Manufacturers
ADAS	Advanced Driver Assistance Systems
ATIS	Advanced Traveller Information Systems
BBC	British Broadcasting Company
CD	Compact Disc
CDS	Crashworthiness Data System
CHI	Computer-Human Interaction
DE	Destination Entry
DSS	Driver Support Systems
DTI	Department of Trade and Industry
EPSRC	Engineering and Physical Sciences Research Council
ESoP	European Statement of Principles
ESRI	Ergonomics and Safety Research Institute
GOMS	Goals, Operators, Methods and Selection Rules
GPS	Global Positioning System
GUI	Graphical User Interface
HCI	Human-Computer Interaction
HGV	Heavy Goods Vehicle
HMI	Human-Machine Interfaces
HVAC	Heating, Ventilation and Air-Conditioning
HTML	Hyper-Text Mark-up Language
ICE	In-Car Entertainment
ISO	International Standards Organisation
ITS	Intelligent Transport Systems
IVIS	In-Vehicle Information Systems
KLM	Keystrok-Level Model
LCT	Lane-Change Test
NASS	National Accident Sampling System
NHTSA	National Highway Traffic Safety Administration
PDA	Personal Digital Assistant
PDT	Peripheral Detection Task

PLATO	Portable Liquid-crystal Apparatus for Tachistoscopic Occlusion
POI	Point of Interest
R	Resumability ratio
Rf	Reach far
RoSPA	Royal Society for the Prevention of Accidents
SAE	Society of Automotive Engineers
SD	Standard Deviation
SDLP	Standard Deviation of Lane Position
SUV	Sports Utility Vehicle
TICS	Transport Information and Control Systems
TSOT	Total Shutter Open Time
Ttclose	Total shutter closed time
TTT	Total Task Time
UK	United Kingdom of Great Britain and Northern Ireland
USA	United States of America
VC	Volume Change

Preface

The research presented within this thesis was principally conducted at the University of Nottingham, between 2003 and 2007. The University of Nottingham and the Transport Research Laboratory (TRL) jointly funded the research.

1 Introduction

This thesis addresses a growing concern for Human-Computer Interaction (HCI) and human factors researchers in the transportation sector; namely, how to evaluate the potential for new in-vehicle information systems (IVIS) to dangerously distract drivers from fundamental, safety-critical driving tasks. The approach combines theoretical knowledge and assumptions with empirical research, ultimately resulting in the proposal of an economical, practical method for addressing this issue.

1.1 Cars and Technology

The task of driving is recognised to comprise of three hierarchical levels: strategic (e.g. overall journey decisions), tactical (e.g. lane choices, route following), and control (e.g. braking, steering) (Michon, 1993). Novel computing and communications technologies, such as navigation systems, adaptive cruise control and collision avoidance, aim to support drivers across all levels of the driving task, and hence improve the safety, efficiency and comfort/pleasure of the driving experience (Bishop, 2005). Indeed, it has been suggested that there are no areas, or vehicle processes, where technology cannot be applied to improve the driving experience, and as such, technology has become an important factor in the marketing of new vehicles (Walker et al., 2001).

Novel in-vehicle systems are referred to within the literature using a wide range of terms (and their associated acronyms), at various levels of specificity. Examples include: driver information systems (Green, 1999a), Advanced

Driver Assistance Systems (ADAS) (Zwaneveld et al., 1999), Transport Information and Control Systems (TICS) (Parkes, 1997), Advanced Traveller Information Systems (ATIS) (Mollenhauer et al., 1997), and, Driver Support Systems (DSS) (Michon, 1993). In the context of this thesis the broad term In-Vehicle Information Systems (IVIS) is used to signify the range of information and communication technologies that are being introduced into the vehicle environment (Young, Regan and Hammer, 2003).

In general, there are three types of novel computing systems that affect the safety, efficiency and comfort of the driving experience. These are illustrated in Figure 1.1 and include: systems that fundamentally change the driving task (e.g. adaptive cruise control, lane keeping assistance); systems that provide information or services relevant to primary driving tasks (e.g. navigation systems, travel information, vision enhancement); and, systems that provide information or services that are non-driving related (e.g. email, internet) (Galer, 1994). These latter two categories cover the systems most commonly referred to as IVIS. It is the interface-design of these types of system with which this thesis is primarily concerned.

An analysis of research concerning IVIS user-interfaces indicates there are three core questions that have been addressed (Noy, 1997, Burnett et al., 2004):

1. What are the impacts (in terms of safety, efficiency, comfort, etc.) of IVIS on drivers, passengers, other road users, and for the traffic/transport system as a whole?

2. Is there potential for novel user-interfaces in a driving context, for example, speech recognition, head-up displays, gesture-driven interfaces, handwriting recognition, and so on?
3. What methods and measures are appropriate for use in the design and evaluation of IVIS?

This thesis addresses the third question in particular, seeking to develop a cost-effective method that is valid and reliable for assessing the visual demand of an IVIS user-interface at early stages in the design process, as part of a driver-centred design approach.

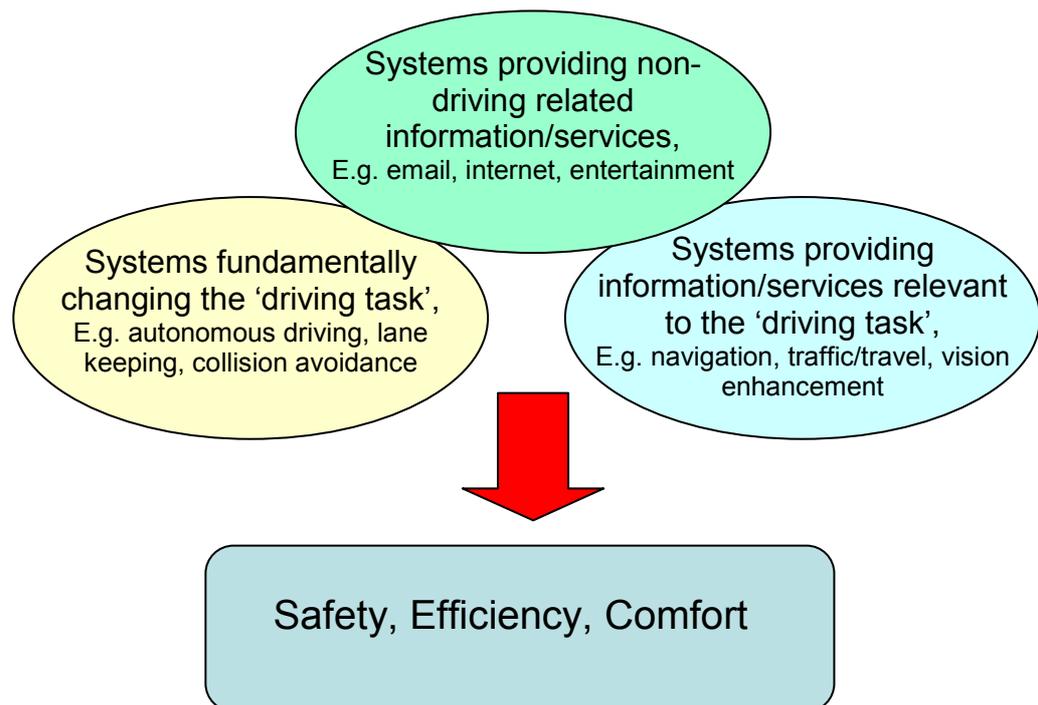


Figure 1.1 The impact of technology on the driving experience (based on Galer, 1994)

Research into the safe design of IVIS interfaces is timely given the systems' increasing prominence, primarily with the rapid rise in the popularity of satellite navigation systems. Built-in navigation systems have been available

in luxury vehicles as expensive options for some time now; however, there has been a significant growth in sales of satellite navigation systems following the introduction of affordable, nomadic systems to the market. In the UK it is reported that 9% of drivers have satellite navigation equipment within their vehicles (Times, 2006a). Furthermore, the boundaries between satellite navigation systems and personal computing devices have become blurred with the introduction of Personal Digital Assistants (PDAs) that provide satellite navigation functionality. Similarly, BluetoothTM functionality enables nomadic navigation systems to act as hands-free phone equipment. As technology and infrastructure develops, drivers are likely to be able to access a range of more advanced functions, for example email and the Internet, whilst on the move (Stevens and Minton, 2001). In addition, car manufacturers are under pressure to offer extra functionality within their vehicles in order to remain competitive.

HCI specialists are concerned that new forms of technology accessible to drivers could become significant sources of distraction and may have an impact on road accidents. Whilst sources of distraction are not limited to technology, a burden of responsibility for minimising the distraction potential of in-car systems lies with the manufacturer. A well-designed system will consider not only the ease of interaction with the device but also the effects that engagement with the interface may have on driver behaviour and performance (Salvucci, 2001). Designers of IVIS are faced with an atypical situation in user-interface design, in that the interface to be designed is 'secondary or peripheral to some performance-critical primary task' (Salvucci, 2001: p. 86). Consequently, designers have a responsibility to consider not just the ease of interaction with

the interface, but the effects that use of the interface may have on driving performance and behaviour (Salvucci, 2001).

The implications of a driver engaging in a secondary task that has adverse effects on driving performance are severe and potentially fatal. It has been suggested that as IVIS become more commonplace there may be an increase in the number of distraction-related accidents, particularly if interfaces are poorly designed (Green, 2000, Stevens and Minton, 2001). Furthermore, a concern is that some IVIS tasks may be highly prioritised by drivers, e.g. those associated with navigation, as they can be viewed as supporting performance of primary driving goals (i.e. successfully reaching a destination). Consequently, drivers may engage in such tasks to the detriment of their driving performance (Cnossen et al., 2004).

1.2 *Distraction and accident rates*

The threat of injury through motor-vehicle accidents is impossible to over-emphasise. In the majority of countries where automobile use is widespread road accidents are one of the major causes of injury, with motor-vehicle-related injury 'estimated as the fourth largest cause of disability-adjusted life years lost in prosperous countries in 1999' (Lam, 2002, p. 412). In the UK, in the year 2005, 3,201 people were killed and 28,954 seriously injured because of automobile accidents (DfT, 2007). There can be multiple causes of a particular accident, with three categories identified as human factors, environmental factors and vehicle factors (Treat et al., 1979, Stutts et al., 2001, Lai, 2005). Of these, human factors are by far the most frequently implicated, with some

involvement in over 90% of accidents (Treat et al., 1979, Stutts et al., 2001, Lai, 2005). Recent statistics from the UK reveal that failure to look properly was the most regularly reported contributory factor, involved in 32% of all injury accidents reported in 2005 (Times, 2006b). In contrast, excessive or inappropriate speeding, which has traditionally been the focus of safe driving government initiatives, was only reported as a contributory factor in 15% of all injury accidents, and 26% of all fatal accidents (Times, 2006b).

Driver distraction is a well-established causal factor in road accidents (Wang et al., 1996, Green, 2000, Stevens and Minton, 2001, Stutts et al., 2001). The precise proportion of accidents attributed to distraction varies between studies, ranging from 2% to 25% and beyond. This variation is due, in part, to inconsistencies in the way accident causes are categorised, and in the types of accident sampled. It can sometimes be difficult to categorise accidents, particularly if there are no witnesses, and especially so in the case of fatal or serious accidents (Stevens and Minton, 2001). Further, studies rarely consider the same data sources, and data sampling can vary substantially between studies.

A concern is that poorly designed IVIS may introduce several dangerously distracting tasks into the driving environment. Increases in the number of vehicles equipped with information and communication technologies could potentially have an impact on the role distraction plays in automobile accidents, or at least on the distribution of distraction sources within the broader inattention category (Stevens and Minton, 2001). A key concern is

that there is no comprehensive definition of driver distraction. Consequently, understanding of the term is vague, and may be confused with the broader category of inattention. This can lead to different estimates of the role distraction plays in road accidents. A working definition of driver distraction to be used in this thesis is as follows: ¹

Driver distraction is:

- *Delay by the driver in the recognition of information necessary to safely maintain the lateral and longitudinal control of the vehicle (the driving task)*
- *Due to some event, activity, object or person, within or outside the vehicle,*
- *That compels or tends to induce the driver's shifting attention away from fundamental driving tasks,*
- *By compromising the driver's auditory, biomechanical, cognitive or visual faculties, or combinations thereof.*

As made explicit in the definition, four inter-related subcategories of distraction have been identified: visual distraction; auditory distraction; biomechanical distraction and cognitive distraction (Young et al., 2003). It has been suggested that IVIS are a potentially worrying source of visual distraction in particular as interfaces for such systems are highly visually orientated, having taken design cues from the desktop computing paradigm (e.g. use of menus, scrolling lists, etc.) (Burnett et al., 2004). Consequently, this thesis is concerned with visual distraction in particular. Whilst other distraction forms are worthy of research it would appear that visual distraction should be given priority. For example, the effects of cognitive distraction, when compared with visual distraction, have been described as very subtle, accounting for much less

¹ The reasoning behind the development of this definition is discussed in detail in Chapter 2

of the overall variance in driving performance (Angell et al., 2006). Furthermore, a longitudinal study of driver behaviour found that visual distraction accounted for the majority of distraction related accidents (Neale et al., 2005, Klauer et al., 2006).

Consideration of the nature of visual distraction suggests it is comprised of two factors: visual demand and willingness to engage. Visual demand, defined as ‘the degree of visual activity required to extract information from an object to perform a specific task’ (ISO, 2005, p. 14), is a property of the system or distraction source. Willingness to engage, is more a property of the driver, and is influenced by several factors, including the priority of the secondary activity, driver experience and the current driving environment. Willingness to engage is therefore difficult to quantify. Instead, research generally focuses on system properties such as visual demand – i.e. the visual distraction afforded by a system when a driver engages in a secondary task – since it is more likely that this can be significantly reduced through design. Whilst design may play a role in minimising willingness to engage, other necessary approaches include driver training and public awareness campaigns. It is not uncommon for the terms visual distraction and visual demand to be used interchangeably, however the distinction is necessary. Unless an assessment measures both visual demand and willingness to engage it is not a true measure of visual distraction.

1.3 HCI methods and driver distraction

A user-centred design approach requires ways of understanding users’ requirements and assessing the appropriateness of alternative user-interfaces

and, in this respect, HCI offers a diversity of methods (Dix et al., 2003). For IVIS design, what is required is a *driver-centred* approach, which considers the specific user needs for systems that are secondary or peripheral to the performance-critical main task of driving (Salvucci et al., 2005). Such an approach may make use of many of the same global techniques familiar within HCI research, but also requires its own methods (or, at least, specific variations of generic methods), which account for the complex, safety-critical characteristics of the driving task. Qualitative techniques may include questionnaires, interviews, guidelines and checklists (Stevens et al., 2002), whilst quantitative techniques may focus on generic concepts such as task times (Green, 1999a), or driving specific measures obtained through simulator or road based studies (Young et al., 2003).

A range of HCI methods are available for assessing driver distraction (for further information, see Chapter 3). Generally, the field has shown a trend for evaluative HCI methods as opposed to generative HCI (e.g. Thimbleby, 1984, Thimbleby et al., 2001). Both qualitative and quantitative methods have been explored, with the latter category further delineated as user-trial or non-user trial based. Three methods are of particular relevance to the later work of this thesis: the 15-second rule, the occlusion technique, and the Keystroke Level Model (KLM). The 15-second rule broadly states that tasks that can be completed in 15 seconds or less in a stationary vehicle (static task time) can be safely performed whilst driving (Green, 1999a). The occlusion technique examines task performance under interrupted vision conditions (occlusion), to assess the visual demand of IVIS (ISO, 2007). The KLM is a modelling

technique that allows for task-time predictions to be made through decomposition of a task into its most basic actions (Card, Moran and Newell, 1983).

A general limitation of many quantitative techniques is that assessment is often made through data gathered from user trials, with a focus on measures of human performance. A practical concern is that such trials require robust prototypes and can, therefore, only be implemented in later design stages. At this point, design changes may prove too costly to apply (Green, 2003, Salvucci et al., 2005, Lansdown et al., 2004). In practice, designers may only have one option; features that do not pass the assessment criteria are likely to be ‘locked-out’ from the driver whilst a vehicle is in motion, instead of making efforts to refine the design such that the task would be acceptable. This situation is not ideal, as it means drivers may be prevented from engaging with systems due to poor interface design, rather than inherently distracting task characteristics (Burnett et al., 2004). There is a need for quantitative methods that can be applied in the earlier stages of the design cycle, for it is at this stage in product development that meaningful changes to an interface design are more practical (Green, 2003, Salvucci et al., 2005).

Common thinking suggests that a viable evaluation method should conform to three principles: validity, reliability, and economy (see Figure 1.2). Validity is a complicated concept, with several types discussed in human factors research (Sanders and McCormick, 1992). In the context of this research, we are predominantly concerned with *face validity*, ‘the extent to which a measure

looks as though it measures what is intended' (Sanders and McCormick, 1992, p. 38), and *construct validity*, 'the extent to which a measure is really tapping the underlying "construct" of interest' (Sanders and McCormick, 1992, p.38). For example, the occlusion technique yields a measure known as total shutter open time (TSOT). The extent to which TSOT relates to total eyes off road time when conducting in-vehicle tasks represents the measure's face validity, whilst the extent TSOT tells us about visual distraction is its construct validity. Reliability has been defined as 'the consistency or stability of the measures of a variable over time or across representative samples' (Sanders and McCormick, 1992, p. 37). For the purposes of this research, reliability refers to the likelihood that a reassessment of an in-vehicle device using a similar sample of participants, under identical conditions, will produce results in keeping with those of a previous study. Finally, economy refers to several factors surrounding a method that relate to its economic cost. These include, the fiscal costs of carrying out the method (e.g. reimbursement of participants, hire of equipment, investigator salaries, and so on), and the time and financial costs associated with a method when used in an iterative design process. Amongst others, Green (2003) and Salvucci (2005) have argued that many existing methods for assessing in-car systems are not useful from an economic viewpoint, since the costs associated with redesign based on their results are too severe.

It is suggested here that no current methods satisfy all three conditions of validity, reliability and economy, principally due to the reliance on performance-based user-trials (for more detail see Chapter 3). Consequently,

the experimental work of this thesis demonstrates the development of a technique that addresses this concern. An extended KLM technique is presented, which seeks to make predictions of the metrics normally generated from measures of performance taken during occlusion user-trials, thus realising the benefits of both the KLM and the occlusion technique.

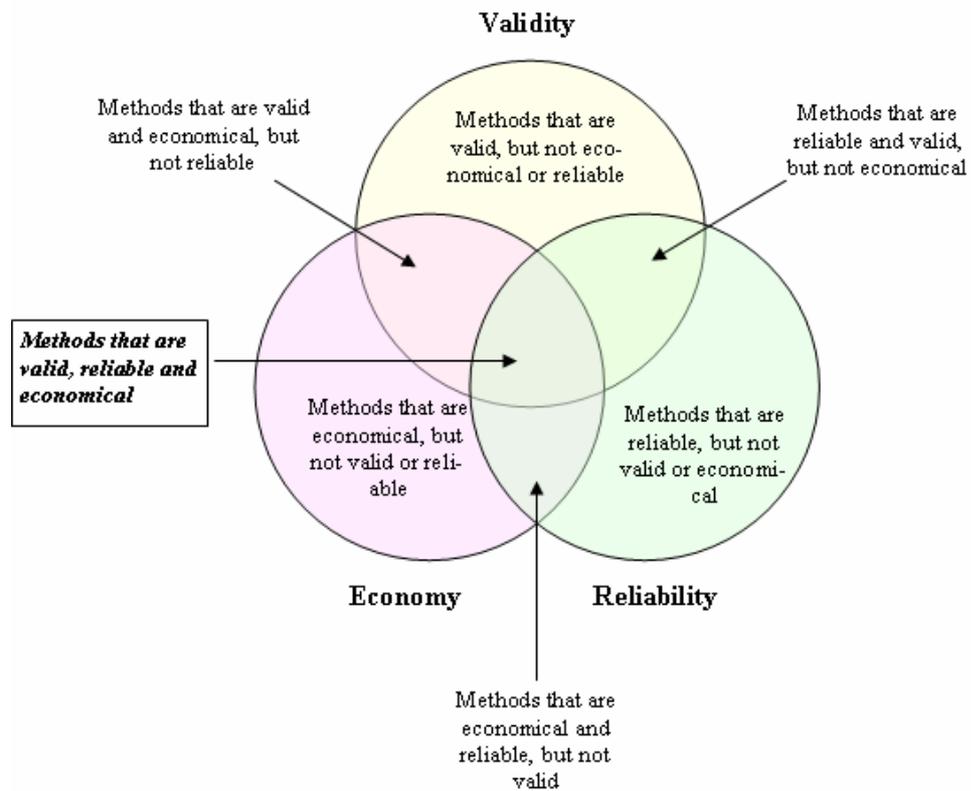


Figure 1.2 Categorisation framework for evaluation methods

1.4 Aims and Objectives

The overall aim of the research is as follows:

To develop a method for assessing the visual demand afforded by IVIS that is valid, economical and reliable.

In order to meet this aim, three objectives are identified.

1. To investigate the issue of driver distraction, clarify the definition of the term and gauge its importance as a factor in accident causation
2. To assess, through empirical assessments and available literature, the range of methods currently available for measuring distraction potential, focusing on their validity, reliability and economy
3. To develop a new, economical method, or expand/adapt existing method(s), for measuring distraction potential, and test its validity and reliability

1.5 Research Contribution

The research presented in this thesis contributes to knowledge in several ways.

Including:

1. The notion of driver distraction is investigated. The discovery of a variety of different definitions in use within the research community leads to a comprehensive definition that is proposed in *Chapter 2* and which was presented to the transport research community at the ITS World Congress in 2005 (Pettitt et al., 2005)
2. The occlusion technique is investigated in *Chapter 4*. The study presents a rare comparison between results of an occlusion study and a road based evaluation of in-vehicle systems, with a prototype system, and, to the authors' knowledge was the first published study of the technique as applied to the emerging ISO standard, and is thus seen as an assessment of the method's practicality. The findings also demonstrated previously unpublished insight into the exact nature of the R metric (Pettitt et al., 2006a).

3. The research demonstrates the development of a new method for evaluating the visual demand of in-car interfaces. Starting with an examination of the keystroke level model (KLM) in an in-car context, the research shows how this technique can be adapted to predict the measures of TSOT and R normally obtained through occlusion user-trials. Since the extended KLM can be applied in less advanced design stages it is argued that it is more economical than traditional occlusion studies, and of more value in the iterative development of in-car devices. Furthermore, the technique is considered useful beyond the context of driver distraction, due to the proliferation of mobile devices in development where techniques to minimise visual demand may be of relevance (Baber, 2001).

1.6 Structure of the thesis

The thesis structure is illustrated in Figure 1.3. Following this introduction, two literature review chapters are presented. *Chapter 2* considers the prevalence of driver distraction as a causal factor in accident statistics, the various definitions of driver distraction that exist, and the formation of a comprehensive definition. *Chapter 3* discusses the range of HCI methods that have been used to evaluate IVIS in terms of distraction, and gives particular focus to the occlusion technique and the KLM. The chapter concludes with a discussion of potential opportunities for research.

The literature review chapters are followed by five studies, which demonstrate the evolution of the extended KLM technique. Table 1.1 presents summary

information for the studies. *Chapter 4* details an assessment of the validity of the occlusion technique. Two tasks on two IVIS are assessed through occlusion trials, as part of a road-based evaluation study. Data was collected by the author, in collaboration with staff from the Ergonomics and Safety Research Institute (ESRI), at Loughborough University, as part of a wider evaluation of a novel IVIS under the Department of Trade and Industry (DTI) and Engineering and Physical Sciences Research Council (EPSRC) funded Foresight Vehicle BIONIC project. The author was involved in the overall study design and assisted in data collection, with chief responsibility for the occlusion trials. Trials took place in the summer of 2004. The author completed all the video analysis for this study, and was solely responsible for the statistical analysis presented in this thesis. The study outlined in this chapter was published in the Institute of Engineering and Technology Journal on Intelligent Transport Systems (ITS), which was selected as featured article for February/March 2007 (Pettitt et al., 2006a).

Table 1.1 Study details

Study	Purpose	Data collection	Location	Date
1	Assessment of occlusion technique	The author, Terry Harrison and Steven Bayer of ESRI	ESRI, Loughborough University	Spring/ Summer 2004
2	Assessment of the KLM in vehicle environment	Darius Karbassioun, University of Nottingham	University of Nottingham	Summer 2004
3	Feasibility of extended KLM	As per Study 1 (occlusion data), the author (extended KLM)	ESRI and University of Nottingham	Early 2005
4	Validity of extended KLM	The author	University of Nottingham	Spring 2006

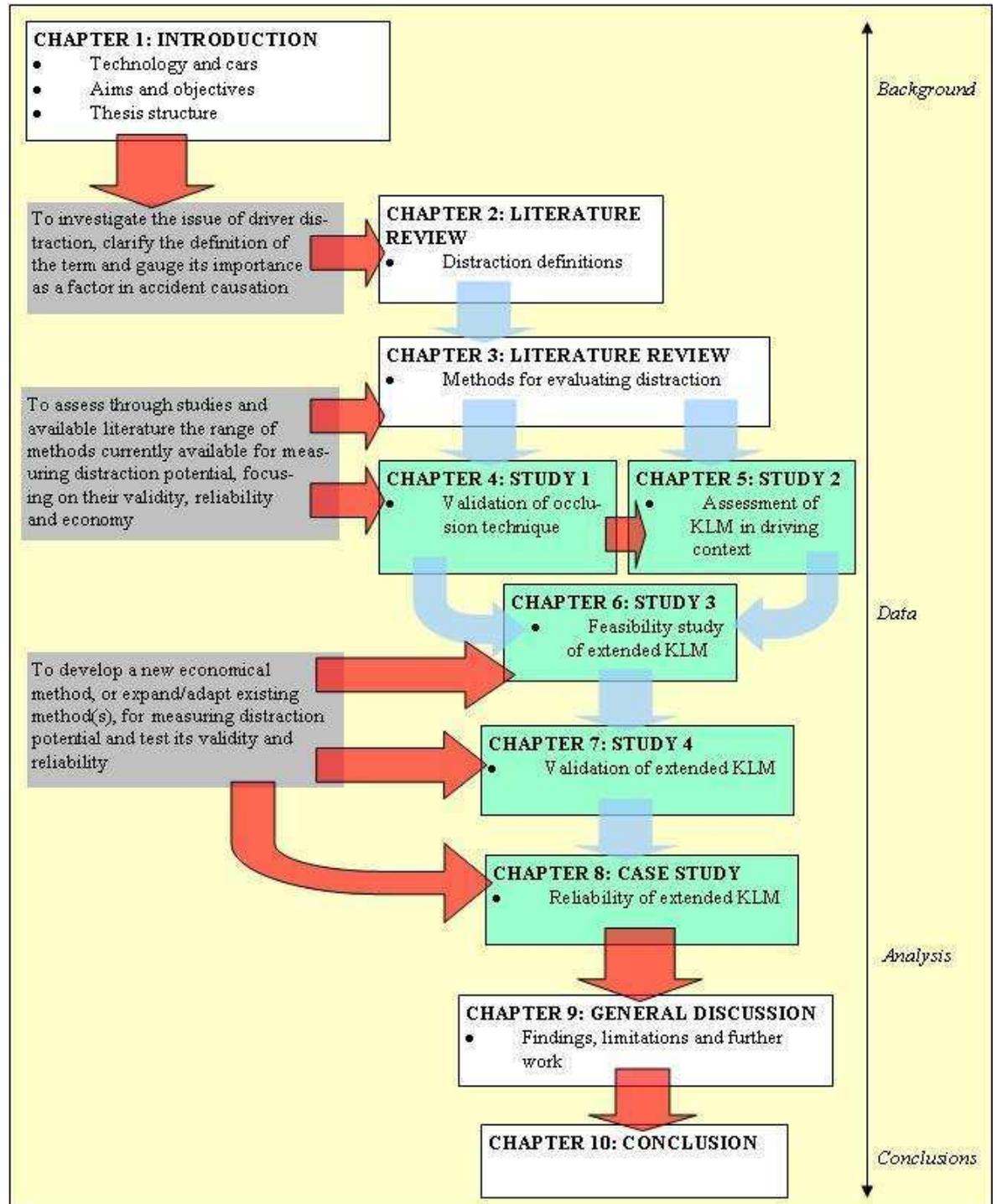


Figure 1.3 Structure of the thesis

The second study, in *Chapter 5*, is an assessment of the traditional KLM within the vehicle environment. The results of a user-trial assessment of performance times for a variety of tasks on in-car entertainment (ICE) systems are compared with task time predictions made through a KLM evaluation. The data

collection for this study was performed by an MSc student, Darius Karbassioun, at the University of Nottingham in the summer of 2004 (Karbassioun, 2004). The author of this thesis provided guidance on the study design (e.g. task selection, methodology), checked and verified the keystroke models, and performed the analysis presented in this chapter. In particular, the discussion of results presented here follows an independent analysis of the data on the part of the thesis author. For example, in his study Karbassioun did not address prediction error rates, which influence the subsequent conclusions made regarding the use of the KLM in the driving context as presented in Chapter 5. Parts of this chapter appear in Pettitt, Burnett and Karbassioun (2006), and selected results were presented at the Ergonomics Society Annual Conference 2006, in Cambridge, UK (Pettitt et al., 2006b).

With the validity of both the occlusion technique and the KLM assured, *Chapter 6* examines the feasibility of modelling performance of tasks under the interrupted vision conditions of occlusion trials using KLM techniques. The data from Chapter 4 is revisited and compared with predictions of occlusion metrics made according to an extended KLM procedure, which is outlined in the chapter. This is a feasibility study, which aims to determine if there is merit in such an approach. Parts of this chapter appear in Pettitt, Burnett and Stevens (2006), and results were presented at the World Congress on Intelligent Transport Systems (ITS) held in London, in October 2006 (Pettitt et al., 2006c).

The extended KLM method is re-assessed in *Chapter 7*, with a wider range of tasks and data gathered from new occlusion trials. Here, the author was solely responsible for study design, data collection, task modelling and analysis. The results of this study are reported in Pettitt, Burnett and Stevens (2007), and were presented at the Computer-Human Interaction (CHI) Conference in San Jose, USA, in 2007 (Pettitt et al., 2007). Thus, Chapters 6 and 7 establish the validity of the extended KLM.

Chapter 8 presents a case study, which investigates the reliability of the extended KLM. A modeller with expertise in HCI, familiar with the KLM and the theory of the extended KLM (but no practical experience of the method), made predictions of occlusion measures for two IVIS tasks. These predictions are compared with the results of an independent occlusion study conducted at the Transport Research Laboratory (TRL), as part of a project evaluating the technique for the UK Department for Transport. In addition to assessing reliability, the results in this chapter allow some comment to be made regarding the economy of the extended KLM technique.

Chapter 9 presents a general discussion of the key findings of the research. This includes an overview of the extended KLM in terms of its validity, reliability and economy. Within the chapter, areas for future work are identified and discussed. In addition, the chapter considers the contribution to research made in the thesis, including the relevance of the research to HCI in general, use of the extended KLM in other contexts, and any implications for

IVIS design. Finally, *Chapter 10* summarises the main conclusions of the thesis.

2 Literature review: driver distraction definitions and accident statistics

Whilst it is generally agreed that driver distraction is an issue of concern for road safety (Young et al., 2003), it is necessary to clarify the precise issue under investigation. This literature review chapter examines the various definitions of driver distraction used in previous research, in order to fully understand the concept, ultimately leading to a synthesised, comprehensive definition. The latter sections of the chapter examine driver distraction in relation to available accident statistics, considering various types of research that seek to establish a link between distraction and accidents.

2.1 Driver distraction: current definitions

Investigations of the effects of distraction on driving have employed varying definitions. It is not uncommon, though, for studies to make no specific definition of the term; an official document such as the European Statement of Principles (ESoP) on in-vehicle Human-Machine Interfaces (HMI), for example, does not clearly define what is meant by driver distraction (Janssen, 2000). This could be related to the common, everyday nature of the word ‘distraction’. Defined in the Oxford dictionary as ‘*something that distracts the attention and prevents concentration*’ (Pollard, 1994), distraction, in the common sense, is a somewhat abstract term, lacking the precision necessary for scientific purposes. In effect, the generally understood meaning of distraction does not cater for the complexities of the term when related to driving.

Definitions currently used in the literature come from a variety of sources, ranging from official bodies or organisations, to the less formal. A review of definitions is presented in Table 2.1. Comparison between definitions highlights a struggle between brevity (e.g. ISO, 2004) and completeness (e.g. Young et al., 2003). What becomes apparent, however, is that whilst most definitions satisfactorily address one or several aspects of the process, there is no single, comprehensive definition.

Comparisons of the available definitions in the literature suggest that a comprehensive definition of driver distraction must take into account the following key components:

- The difference between distraction and inattention;
- Recognition that distraction can be internal or external to the vehicle;
- The effect of distraction on the driving task.

In addition to the definitions presented in Table 2.1, distraction has been discussed within the literature in terms of four distinct, but not mutually exclusive, types: auditory distraction, biomechanical distraction, cognitive distraction and visual distraction (Ranney et al., 2000, Young et al., 2003). These types should therefore form a part of any comprehensive definition. These components are discussed in detail below.

Table 2.1 Definitions of distraction

Organisation/author	Definition of...	Definition	Comment	Source
ISO	Distraction	Attention given to a non-driving related activity, typically to the detriment of driving performance	Concise, and considers the effect of distraction on the driving task.	ISO, 2004, p. 10
ISO	Visual distraction	Diversion of visual attention from the road scene	Brief and of limited use, unless in conjunction with a wider definition	ISO, 2004, p. 10
Indiana Tri-Level Study of the Causes of Traffic Accidents	Internal distraction	Attention given to a competing event, activity, or object inside the vehicle	Specific to internal distraction. Effective use of the word 'competing' – suggests conflicting attentional demands of the driving task and the distraction source.	Wang et al., 1996, p. 2
American Automobile Association Foundation for Traffic Safety (AAAFTS)	Driver distraction	Delay in the recognition of information needed to safely accomplish the driving task because some event, activity, object or person within or outside the vehicle compelled or tended to induce the driver's shifting attention away from the driving task	Comprehensive. Acknowledges multiple sources of distraction. Recognises the demand placed on attention resources and potential for task sharing. Doesn't mention distraction types.	Young et al., 2003, p. 2
Ranney, Mazzae, Garrott and Goodman Lam	Driver distraction Driver distraction	Any activity that takes a driver's attention away from the task of driving Any movement, activity, and/or happening initiated by the driver or someone else inside or outside the vehicle that might have caused a diversion of the driver's attention	Concise, but lacks specificity. Use of the word 'might' is problematic. Just because an activity 'might' be a source of distraction does not make it so, there needs to be an impact on driving performance.	Ranney et al., 2000, p. 1 Lam, 2002, p. 413
Lam	Internal distraction	Any activities, actioned by the driver or passenger(s), which might cause a diversion of the driver from his/her driving task	Similar to above.	Lam, 2002, p. 413
Janssen	Driver distraction	The capture of the driver's attention by information that is irrelevant to the driving situation to a degree where insufficient attention is left for the primary task	'Insufficient attention' a good description – distraction does not necessarily take <i>all</i> attention from driving task. Not sure all distraction sources completely fall into category of irrelevant information.	Janssen, 2000, p. 1 - 2
Stevens and Minton	Driver distraction	A physical event or action or condition in or on the vehicle, which diverts the driver's attention from the driving task	Reasonably concise term, although no mention of distraction types, what is meant by the driving task and the effect of distraction	Stevens and Minton, 2001
100-car study	Driver distraction	When a driver has chosen to engage in a secondary task that is not necessary to perform the primary task	Suffers from similar problems as Janssen and Lam. Assumes all secondary tasks detrimental to driving.	Klauer et al., 2006

2.1.1 Distraction and inattention

Attention has been defined as our ability to focus on a task, and has been linked to the key human factors concept of situational awareness – which, in simple terms can be described as knowing what is going on around you (Endsley, 1988, Endsley and Garland, 2000, Tasca, 2005).² Good situational awareness allows drivers to determine the appropriate level of attention required in a given driving scenario (Tasca, 2005). Three levels of attention have been described: selective attention, divided attention and automatic attention (Tasca, 2005). The dynamic nature of the driving environment means that the appropriate level of attention to ensure safe driving is constantly changing, and is determined by the demands of the road environment (Tasca, 2005). For example, making a right-hand turn on a busy urban street with multiple lanes is an instance where selective attention is required; the driver attends fully to the driving task and ignores irrelevant information or tasks (Tasca, 2005). In less demanding driving scenarios, it may be possible to engage in less important tasks, whilst maintaining an acceptable level of driving performance.

Drivers may feel they are able to divide their attention between primary driving tasks and secondary tasks; however this perception may be misinformed. The success or failure of a person's ability to divide attention between multiple tasks has been linked to four factors: resource demand, structure, similarity, and resource allocation/task management (Wickens et al., 2004). Since many in-vehicle tasks draw on similar mental resources to the driving task, it is

² A more precise definition of situational awareness states the phenomena is 'the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future' (Endsley, 1988, p. 97).

unlikely that drivers will be able to successfully share their attention between the two tasks (Wickens et al., 2004). Consequently, in order to multi-task, a person must constantly shift attention back and forth between tasks as part of a time-sharing strategy (Tasca, 2005, Wickens et al., 2004). When attention is misallocated between driving and secondary tasks, or when drivers misjudge the attentional demands of the road environment, there is a risk of breaking down the complex information-processing loop involved in maintaining driving performance, potentially resulting in an accident (Lai, 2005). The attentional demands of driving require drivers to perform complex information processing tasks in split-second periods of time (Lai, 2005, Tasca, 2005); ‘The quantity and quality of information available to the driver are a function of his/her level of attention’ (Tasca, 2005: p. 7). Furthermore, the ability to successfully share attention between tasks has been shown to degrade with age (McKnight and McKnight, 1993, Lam, 2002, Young et al., 2003). With an ageing population, and growing proliferation of distraction sources within vehicles, it is important to be aware of this fact.

Driver distraction has been viewed as *one* part of a driver inattention construct (Neale et al., 2005). Here, distraction is viewed as *secondary task engagement*, with the other elements in the construct being: *fatigue*, *driving-related inattention to the forward roadway*, and *non-specific eye glances* (Neale et al., 2005). Driving-related inattention to the forward roadway refers to times when drivers are checking mirrors or blind spots; attention is still focused on driving tasks, but not on the fundamental, or critical, aspects, i.e. monitoring the

forward roadway environment (Neale et al., 2005). Within this construct, it is unclear where daydreaming should be categorised.

Consequently, definitions of distraction must consider the presence of an event or occurrence that causes a driver to allocate attention, which might otherwise be focussed on the driving task, to a separate activity. Where this has been achieved, caveats used have discussed an ‘event, activity, object or person within or outside the vehicle [that] compelled or tended to induce the driver’s shifting attention away from the driving task’ (Young et al., 2003, p. 2), or physical events, actions or conditions, in or on the vehicle that divert attention from driving (Stevens and Minton, 2001).

Distraction is a specific form of inattention, when the drivers’ attention level has been misallocated between fundamental driving tasks, and competing activities. Whilst there are other forms, and causes of inattention, e.g. daydreaming, it is the presence of a triggering event, object or activity that distinguishes distraction (Young et al., 2003). The key consideration is that the result of distraction is inattentive driving; however, inattention is not always caused by distraction (Beirness et al., 2002, Tasca, 2005).

2.1.2 Internal and external distraction

Sources of distraction can be both internal and external to the vehicle. Where a definition encompasses both internal and external distraction, it takes into account distractions that are initiated by the driver and also those that are non-driver initiated, examples of which include ‘acute situations that demand a

quick response from the driver' (Lam, 2002, p. 416). Internal distractions can be categorised as both driver initiated, e.g. making a mobile phone call, or non-driver initiated, e.g. the unpredictable actions of a passenger. It has been stated that all forms of external distraction could be said to be non-driver initiated, e.g. the unpredictable behaviour of a drunk pedestrian in the roadway (Lam, 2002). However, some external distractions may be less easily categorised. Drivers may potentially be distracted from fundamental driving tasks when scanning road or street signs, for example. Such instances could be seen as driver initiated distractions. In a review of work-related road accident statistics, it was found that a significant proportion of distraction related accidents occurred when drivers were looking for, or at, landmarks or street names, and not paying attention to the forward roadway (see Appendix A).

The National Highway Traffic Safety Administration (NHTSA) recognises 13 distraction sources as follows (Stutts et al., 2001, Young et al., 2003):

1. Eating or drinking
2. Outside person, object or event
3. Adjusting radio, cassette or CD
4. Other occupants in vehicle
5. Moving object in vehicle
6. Smoking related
7. Talking or listening on mobile phone
8. Dialling mobile phone
9. Using device/object brought into vehicle
10. Using device/controls integral to vehicle

11. Adjusting climate controls

12. Other distraction

13. Unknown distraction

These distraction sources can be categorised in several ways, e.g. technology and non-technology based, internal and external, and so on (Young et al., 2003). All sources other than Category 2, and potentially Categories 12 and 13, are internal distractions. However, whilst many of the sources are fairly specific, e.g. separate categories for talking/listening on a mobile phone (Category 7) and for dialling (Category 8), Category 2, ‘outside person, object or event’, is a fairly general description, and could cover a variety of situations.

It can be difficult to determine the level of specificity required for accident reporting. Whilst it may be felt useful to provide high levels of specificity, such as with the separate categories for mobile phone interactions, in practice this may not prove to be the case. An analysis of crash reports using the above categories found there was no strong case for distinguishing between dialling mobile phone numbers and talking or listening (Stutts et al., 2001).

2.1.3 Types of distraction

In order to better understand the nature of driver distraction, the term has been divided into four distinct, but not mutually exclusive, types that are commonly referred to in the literature (Ranney et al., 2000, Young et al., 2003). These are:

- **Auditory distraction** – ‘occurs when the driver momentarily or continually focuses their attention on sounds or auditory signals rather than on the road environment’ (Young et al., 2003: p. 2);
- **Biomechanical distraction** – this involves movements of the driver’s body away from the standard posture required to perform the physical tasks associated with safe driving (Young et al., 2003);
- **Cognitive distraction** – ‘includes any thoughts that absorb the driver’s attention to the point where they are unable to navigate through the road network safely and their reaction time is reduced’ (Young et al., 2003: p. 2);
- **Visual distraction** – comes in three forms: driver’s visual field is blocked, e.g. by dark window tints, stickers on windscreen; when a driver focuses visual attention on something other than the road ahead, e.g. a navigation system; a loss of visual ‘attentiveness’, also known as ‘looked but did not see’ (Young et al., 2003). For the purposes of the thesis, the term visual distraction generally refers to the second form.

2.1.4 Distracting activities and their relationship to the primary driving task

The International Standards Organisation (ISO) has defined distraction as ‘attention given to a non-driving related activity, typically to the detriment of driving performance’ (ISO, 2007). By emphasising the ‘detriment of driving performance’ the definition demonstrates the effect of a distracting activity on the primary task – driving. Consequently, in order to be classified as distracting, a secondary task must have some measurable impact on driving performance, and a definition of driver distraction should cater for this fact.

A criticism could be levelled at ISO's definition for referring to 'non-driving related' activities. Primary driving tasks include: steering, accelerating, braking, speed choice, lane choice, manoeuvring in traffic, navigation, communicating with other road users, and scanning for hazards (Tasca, 2005). It can be argued that route finding, which may involve distracting activities such as using a navigation system or reading a map, is related to the driving task, as the driver has the goal of travelling from one location to another (Tasca, 2005). Research has suggested that drivers are likely to prioritise those tasks that are deemed of relevance to driving; consequently accepting greater detriments in driving performance in order to ensure successful completion of the secondary task (Cnossen et al., 2004).

Cnossen et al. (2004) hypothesised that drivers prioritise activities that serve main driving task goals over less relevant tasks. In a driving simulator study participants completed an auditory working memory task whilst receiving audio route guidance information and whilst following a schematic paper map route. Participants also drove following audio route guidance and paper maps without the working memory task. Trials took place in two traffic conditions, high and low density. The authors found that participants actively sought to maintain performance in tasks related to driving over irrelevant tasks. When navigating with a map in high-density traffic, participants reduced their driving speed. However, this compensatory behaviour may not be sufficient to regulate crash risk, and could result in other drivers perceiving the distracted driver's performance as erratic. Such compensatory behaviour was not observed when participants drove in busy traffic whilst performing the working

memory task. Performance in the working memory task was also found to suffer in conditions where participants drove whilst navigating with a paper map. It is concluded that this is evidence that drivers prioritise performance in a driving-related task (navigation) over non-driving-related tasks (working memory task). This has practical implications, particularly for the design of navigation systems. Whilst drivers may ignore tasks of less importance to their main driving goal, e.g. making a phone call, they will be highly motivated to interact with a navigation system, which serves the overall driving task goal of reaching their destination (Cnossen et al., 2004).

Distraction sources can be categorised as driver initiated and non-driver initiated (Tasca, 2005). Driver initiated sources include: eating/drinking, grooming, using in-vehicle devices, and so on (Tasca, 2005). Non-driver initiated distractions include external distractions, such as the unexpected behaviour of other road users or pedestrians, and internal distractions, e.g. a moving object within the vehicle. It is possible that some distraction sources can fall into both categories. For example, a ringing mobile phone is a non-driver initiated distraction. However, if the driver chooses to accept the call, the distraction could now be classed as driver initiated. It is important to recognise that, once initiated, drivers tend to be highly motivated to complete secondary tasks (Tasca, 2005).

2.2 Working definition

An analysis of currently employed definitions of driver distraction suggests that lack of common understanding of the term may be a hindrance to research

aims. For the purposes of scientific research, a comprehensive and accurate definition is required. A reasonable, exhaustive definition of driver distraction is comprised of four components: impact, agent, mechanism and type, and is as follows:

Driver distraction:

- Delay by the driver in the recognition of information necessary to safely maintain the lateral and longitudinal control of the vehicle (the driving task) (**Impact**)
- Due to some event, activity, object or person, within or outside the vehicle (**Agent**)
- That compels or tends to induce the driver's shifting attention away from fundamental driving tasks (**Mechanism**)
- By compromising the driver's auditory, biomechanical, cognitive or visual faculties or combinations thereof (**Type**).

This definition was presented to the transport research community at the ITS World Congress 2005, with positive response (Pettitt et al., 2005).

2.3 Distraction as a factor in accident causation

It can be difficult to ascertain the precise role that distraction plays in accident causation. Individual studies have had some success in evaluating the size of the distraction problem, although comparisons can be difficult to make (Wang et al., 1996, Stutts et al., 2001, Stevens and Minton, 2001). This is due to a multitude of factors, including disparity between definitions and inconsistencies in accident reporting (Stevens and Minton, 2001). In-depth analysis of accident causation generally only takes place when there has been a serious injury or fatality (Stevens and Minton, 2001, Stutts et al., 2001).

Studies have been hindered by the superficial nature of police reports not designed to provide scientific determinations of crash causation (Wang et al., 1996). Furthermore, it may be the case that distraction plays a disproportionately small role in fatal and severe accidents in comparison to less severe accidents (Stevens and Minton, 2001). Data taken from severe accidents, e.g. fatal accidents, is likely to show high proportions associated with alcohol, fatigue, speeding and aggressive driving (Stevens and Minton, 2001). In recent statistics from the UK, loss of control was the most frequently reported contributory factor, involved in 35 % of road deaths (Times, 2006b). However, it has been found that distracted drivers are no more or less likely to be fatally or severely injured than attentive drivers (Stutts et al., 2001). In severe accidents it can be difficult to draw firm conclusions on causation, due to memory recall problems or lack of witness evidence in fatal cases (Stevens and Minton, 2001). Often, accident causation can only be inferred through circumstantial evidence (Stevens and Minton, 2001).

An examination of roadway factors suggests that distracted drivers are involved in crashes under different roadway conditions than attentive drivers. For example, 37.1% of distracted drivers were found to be travelling on multilane roadways, compared with 50.4% of attentive drivers (Stutts et al., 2001). Similarly, it has been found that drivers distracted by cell-phones are over-represented in multi-vehicle crashes (Stutts et al., 2003). When evaluating the role of driver distraction in accident causation, it is important to consider what *types* of accidents distracted drivers cause. Such data is useful

for manufacturers designing collision mitigation and safety equipment, and for educating the driver population about risk (Neale et al., 2005).

2.3.1 Case-control studies

To date, research has struggled to quantify the precise risk posed by distraction in general. One approach is to use case-control studies, a familiar method in epidemiological research, which examine the prevalence of a factor, e.g. distraction, in a population, e.g. motor-vehicle collisions, and compare this to the prevalence of that factor within the wider population, e.g. the driving population (Simpson, 2005). Such studies were conducted over 40 years ago concerning alcohol and accident risk, and were instrumental in setting legal limits for alcohol consumption and acceptable driving performance (Simpson, 2005). A major difficulty for carrying out such studies with distraction is quantifying the level of distraction in the general driving population. Common sense suggests that almost all drivers are likely to experience some form of distraction during a journey, although it is likely to be intermittent (Stevens and Minton, 2001). Consequently, a legislative approach to reducing driver distraction would be complicated, difficult to enforce, and may ultimately be unfair.

A recent study in the USA sought to produce the type of information concerning driver distraction normally generated by case-control studies. The 100-car naturalistic driving study, undertaken to provide researchers with a greater understanding of pre-crash causal and contributing factors, took place in Virginia, USA (Neale et al., 2005, Klauer et al., 2006). One hundred

instrumented vehicles, with a sample of 241 primary and secondary drivers, provided a data set covering approximately 2,000,000 miles, and yielding 43,000 hours of data over a 12 to 13 month period. Over the course of the investigation, the experimenters captured data on 82 crashes (of which 69 had sufficient material for analysis), 761 near-crashes, defined as conflict situations requiring rapid, severe, evasive manoeuvres to avoid a crash, and 8,295 incidents, where less severe evasive manoeuvres were required. Crashes were further delineated into four categories:

1. Level I: police-reported air bag deployment and/or injury
2. Level II: police-reported property damage only
3. Level III: non-police-reported property damage only
4. Level IV: non-police-reported low-g physical contact or tire strike
(greater than 10 mph)

The authors determined four categories of driver inattention: secondary task engagement, which would normally be termed driver distraction; fatigue; driving-related inattention to the forward roadway, where drivers were attending to mirrors or blind spots; and, non-specific eye glances, where drivers glance, momentarily, from the forward roadway, but at no discernable object or person.

Results showed that 78% of crashes and 65% of near-crashes were coded with one of these inattention categories as a contributing factor, with secondary task engagement the greatest source of inattention, accounting for 22% of accidents (Neale et al., 2005, Klauer et al., 2006). Overall, results suggest that engaging in secondary tasks whilst driving increases the risk of having an accident by

two times that of normal baseline driving for simple tasks, and three times for complex tasks (Klauer et al., 2006).

The greatest source of driver distraction was wireless devices, the majority of which were mobile phones. These devices accounted for about 8.7% of all crashes, or near crashes, and normally occurred during conversations (Klauer et al., 2006). The results suggest that dialling tasks on hand-held devices are riskier than talking/listening, however the overall risk of both activities is similar since drivers are likely to talk/listen on such devices for a significantly greater period of time than engaging in dialling tasks (Klauer et al., 2006).

An analysis of eye glance behaviour in the dataset demonstrates that single glances away from the road greater than two seconds significantly increases the risk of being involved in a crash, or near-crash (Klauer et al., 2006). In contrast, glances of less than two seconds do not appear to significantly increase risk beyond baseline driving (Klauer et al., 2006). Furthermore, some glances away from the road were considered important in enhancing safety. However, these glances are associated with driving-related activities, e.g. looking at rear-view mirrors. Overall, driving-related inattention to the forward roadway resulted in reduced crash and near-crash risk. It is suggested this finding demonstrates that drivers scanning the whole driving environment are alert and engaged in the driving task (Klauer et al., 2006).

The 100-car study provides interesting results, and establishes specific crash risk values to a variety of inattention related activities. The dataset is huge, and

is likely to yield results for years to come (Neale et al., 2005). However, certain caveats must be attached to the results. For example, results are specific to the participant population involved in the study, which may not be fully representative (Neale et al., 2005). Participants were drawn from a particular geographic region, with mileage on rural roads and in Sports Utility Vehicles (SUVs) below the national average in the USA (Klauer, 2006). Furthermore, the number of crashes observed over the course of the study is relatively small in comparison with samples gathered from accident causation databases.

2.3.2 Results from around the world

Estimates suggest that between 25% and 50% of police-reported crashes in the USA involve some form of driver inattention (Stutts et al., 2001). The Indiana Tri-Level study of accident causation found internal distraction to be a causal factor in 9% of accidents examined (Treat et al., 1979, Stutts et al., 2001). This research was performed in the mid-1970s, and a number of new potential distraction sources have emerged since it was carried out. Wang et al. (1996) and Stutts et al. (2001) report results from the National Accident Sampling System (NASS) Crashworthiness Data System (CDS). The CDS was developed to obtain detailed information on inattention related accidents. Data is representative of USA passenger vehicle tow-away crashes, including interviews with drivers, inspections of crash scenes and other supporting data sources. Approximately 5,000 randomly selected tow-away crashes are studied per annum. The earlier study found that 7.8% of drivers involved in crashes were distracted and 13.3% of crashes involved distraction (Wang et al., 1996).

For results taken between 1995 and 1999, weighting the data to reflect national results, 12.9% of drivers involved in accidents were identified as distracted, with 8.3% the percentage of vehicles involving a distracted driver (Stutts et al., 2001). However, this is felt to be an underestimate of the percentage of actual crashes involving driver distraction, due to difficulties in determining accident cause and the types of accidents sampled (Stutts et al., 2001).

2.3.3 Results from the UK

The contribution of distraction to automobile accidents in the United Kingdom is difficult to quantify. Three potential sources of information regarding fatal accidents are available – special in-depth investigations, the ‘STATS19’ database of all injury accidents, and, police fatal accident reports (Stevens and Minton, 2001). In-depth investigations are undertaken in special cases only, involving co-operation between hospitals and police forces, and are expensive and restricted in scope (Stevens and Minton, 2001). ‘STATS19’ contains ‘useful but limited data’ (Stevens and Minton, 2001: p. 540). Police fatal accident reports are compiled routinely when a fatality occurs because of an automobile accident, and have long been seen as an important tool in accident research (Stevens and Minton, 2001). Reports for accidents in England and Wales are stored for analysis after all legal requirements have been met, generally about three to six years after an accident takes place. An assessment of accident data stored in these files is not representative of the UK as a whole, as they only contain data from English and Welsh police forces.

In-vehicle distraction is shown to be a contributory factor in 2% of fatal accidents between the years 1985 and 1995 (Stevens and Minton, 2001). The

greatest source of internal distraction is passengers, followed by entertainment systems, and eating and drinking. Further weight may be given to certain sources of distraction than initial analysis would suggest when exposure to risk is taken into consideration (Stevens and Minton, 2001). For example, fewer than half of the accidents examined involved multi-occupant vehicles, magnifying the significance of passengers as a distraction source (Stevens and Minton, 2001).

When making comparisons between studies, caution is advised (Stevens and Minton, 2001). For their study, Stevens and Minton suggest comparison is only possible with the results of Wang et al. (1996). Since Stutts et al. (2001) is based on the same data source as Wang et al. (1996), it is probable that this study may also be comparable. Stevens and Minton's finding that internal distraction accounts for 2% of distraction related accidents is much lower than the figures given in either study based on the CDS. This may be attributed to the elimination of accidents involving vehicles with professional drivers (HGVs, buses, etc.) in the USA studies (Stevens and Minton, 2001). These drivers may be less prone, or less exposed, to distraction; for example, passengers are a less likely source of distraction in the case of HGV drivers (Stevens and Minton, 2001). Furthermore, the data in the CDS is not confined to fatal accidents (Stevens and Minton, 2001). The difficulties involved in determining accident causation, particularly when distraction is a causal factor, in fatal accidents, leads to the assertion that a figure of 2% is likely to be a conservative estimate (Stevens and Minton, 2001).

A more recent UK based study of accident rates and distraction can be found in Appendix A (see also Pettitt et al., 2005). This study used data of work-related road traffic accidents in the Midlands region of the United Kingdom, which includes drivers of a variety of vehicles, including Heavy Goods Vehicles (HGV), buses, and, company or fleet cars. Furthermore, accidents documented within the database include those caused by other road users, with work-related drivers not necessarily at fault. 3.7% of accidents were found to have distraction as a causal factor, a small figure in comparison with some other studies. This may be due in some part to the specific characteristics of the driver population under consideration. For example, passengers, often a significant source of driver distraction, may be under-represented in the sample, given that work-related drivers generally drive alone (e.g. HGV drivers, fleet cars). Furthermore, the data in this study was obtained from police reports. Therefore, similar to Stevens and Minton (2001), there may be an element of underreporting due to drivers not wishing to admit responsibility for an accident. How data is collected in studies of accidents and distraction is a further factor that makes comparison between studies problematic. Independent reports, where drivers are not concerned about incriminating themselves, are likely to provide more honest data.

2.3.4 Sources of distraction

Stutts et al. (2001) report that external distraction – outside persons, objects or events – are the most frequently reported source, at 29.4% of distraction related accidents, followed by making adjustments to the radio/cassette/CD player (11.4%) and other occupants in the vehicle (10.9%). Despite variations in the year-to-year data, these three categories are highlighted as most important

(Stutts et al., 2001). Other distractions, such as moving objects within the vehicle, objects brought into the vehicle, adjusting vehicle or climate controls, eating and drinking, using mobile phones, and smoking account for between 1% and 4% of distraction related accidents each. Two further categories, ‘other’ distracting events and unknown distractions, account for 25.6% and 8.6% of accidents respectively. A summary of these results is presented in Table 2.2.

Table 2.2 Summary of weighted CDS data for driver distraction (adapted from Stutts et al., 2001)

Driver attention status	1995-1999	Driver distraction	1995-1999
Attentive	48.6%	Outside person, object, event	29.4%
		Adjusting radio/cassette/CD	11.4%
Distracted	8.3%	Other occupant	10.9%
		Moving object in vehicle	4.3%
Looked but didn't see	5.4%	Other device/object	2.9%
		Vehicle/climate controls	2.8%
Sleepy or fell asleep	1.8%	Eating, drinking	1.7%
		Using/dialling cell phone	1.5%
Unknown/no driver	35.9%	Smoking related	0.9%
		Other distraction	25.6%
		Unknown distraction	8.6%

Examining in-vehicle distraction, Stevens and Minton (2001) found passengers to be the greatest single source of distraction, followed by interacting with the radio/cassette player and handling food, drink and cigarettes. As noted in Section 2.2.2, it is important to be aware of exposure to a distraction source when evaluating the relative risk (Stevens and Minton, 2001). Whilst passengers, when present, remain in the vehicle for the duration of a journey, it can be assumed that other sources only constitute a distraction at specific times. For example, a car radio is only a distraction when it requires the driver's attention, e.g. for re-tuning (Stevens and Minton, 2001). Furthermore,

whilst passengers are a particularly large source of in-vehicle distraction, their presence in a vehicle is not routine (Stevens and Minton, 2001).

The study of work-related road traffic accidents found that internal distraction sources were more common than external distraction; technology sources accounted for 15% of all distraction-related accidents (see Appendix A). Furthermore, by considering a number of internal and external sources (e.g. map reading, looking for buildings, reading road signs) under a single category of *route finding*, this became the largest source of distraction. This finding emphasises other research, which suggests that drivers may accept greater detriments to their driving ability when secondary tasks serve higher task goals (Cnossen et al., 2004).

2.3.5 The role of technology in distraction-related accidents

Determining the role information and communication technology plays in distraction related accidents is made more complicated by the relative minority of vehicles equipped with such devices (Stevens and Minton, 2001). Consequently researchers are faced with the dilemma of whether to over emphasise the cases of distraction with such technology that do appear in accident statistics, due to their under-representation amongst the driver population. Logically, accidents caused by distraction associated with such devices are likely to be very low in number in comparison to other distraction sources; however, their rarity may imply higher levels of associated risk, due to the comparative minority of vehicles equipped with such devices. Furthermore, crash databases tend to take time to populate. As Stevens and

Minton show, it can be up to six years before data is available. IVIS technologies, such as satellite navigation systems, have only recently become affordable and popular. Therefore, it may be several years before their true impact, in terms of accidents, is made clear.

Distraction arising from interaction with in-vehicle technology is generally termed 'voluntary distraction' or driver initiated distraction; since it is likely the driver chooses to engage in the distracting activity (Stutts et al., 2001, Tasca, 2005). This is dissimilar to involuntary distractions, such as moving objects within the vehicle, or many external distractions. For voluntary distractions, it is likely that drivers will be less willing to engage in activities when driving task demands are high. In an observational study of driver behaviour, it was discovered that a number of voluntary distractions engaged in by drivers took place at times when the vehicle was stationary, for example waiting at traffic lights (Stutts et al., 2003). Clossen et al. (2004) drew similar conclusions, finding that drivers would take advantage of times when the road environment required the vehicle to be stationary to consult paper maps. Therefore, the incidence of such distractions, and the role they play in crashes, may vary 'as a function of roadway, environmental, and vehicle conditions' (Stutts et al., 2001: p. 19). It is argued that this relationship is complex – 'factors that decrease the likelihood of engaging in a behaviour may also increase the likelihood of a crash in the presence of that behaviour' (Stutts et al., 2001, p. 19). The precise interference of secondary tasks on primary driving tasks is dependent on particular factors, namely: timing of the distraction; distraction duration; ease of resumption of primary driving task;

physical location of the distraction; and, the ability of the driver to divide attention (Tasca, 2005).

Due to the lack of data available regarding driver distraction accidents resulting from interaction with in-vehicle technology, research has focused instead on the impact use of such technology has on measures of driving performance, occasionally relating results to crash risk. Comparisons have been made between new in-vehicle technologies, e.g. navigation systems, and existing technologies, e.g. car radios, to establish relative risk (Tijerina et al., 1998). As previously noted, results from the 100-car study suggest that the risk of being involved in a crash, or near crash, is increased by a factor of two for simple secondary tasks, and a factor of three for complex secondary tasks (Klauer et al., 2006).

Navigation systems have come under particular focus of research. Destination entry is generally considered the most complex task associated with such systems (Young et al., 2003). This task has been shown to take substantially longer than dialling a phone number on a mobile phone, or tuning a radio (Tijerina et al., 1998). With some navigation systems, entering a destination whilst driving can take as long as nine minutes (Young et al., 2003). Some systems block out this functionality whilst a vehicle is in motion, although 25% of navigation system users admit to regularly, manually entering destination information whilst driving (Young et al., 2003).

There is a high industry incentive to incorporate more technology in cars (Parkes, 2005). Although route guidance information provided by navigation systems can be a source of distraction, such systems have been shown to offer benefits over traditional way-finding methods such as paper maps, if the information is well presented (Young et al., 2003). Future technology developments may focus on providing drivers with access to email and the Internet, which, although not directly linked to fundamental driving tasks, may help increase drivers feeling of productivity (Stevens and Minton, 2001, Young et al., 2003, Karputer, 2006). Despite design efforts to minimise distraction, serious concerns have been raised regarding the safety of such systems (Young et al., 2003).

Well-designed interfaces, aimed at reducing potential distraction, are viewed as the ideal solution. However, in practice, attempts to produce such interfaces have yielded mixed results. Whilst there is evidence that voice-input or audio-based interfaces are more effective than more traditional, visual/manual designs (Tijerina et al., 1998, Angell et al., 2006), other studies suggest these novel interfaces are not entirely desirable (Young et al., 2003). Lee et al. (2001) demonstrate that using a speech based email interface whilst driving leads to degradations in driving performance. In a simulator-based study, participants' reaction times to a braking lead vehicle were 30% longer than when not using the system. This translated to a 38.5% increase in collisions, and an 80.7% increase in collision velocity (Lee et al., 2001). It is not clear, however, whether the distraction effect is inherent to the email *task* or to the *interface* design, since the authors do not compare their results with an

alternative, visual/manual interface. The distraction may be inherent to the task of interacting with email, regardless of input/output modalities (Young et al., 2003).

Case study: mobile phones

Whilst it can be difficult to ascertain the impact of more novel information and communication technologies on accident causation, due to the delays involved in obtaining data and the relative minority of vehicles equipped with such devices, a large body of research has amassed concerning the use of mobile phones and driving. For example, use of phones whilst driving was the third greatest distraction source in the work-related road accident study (see Appendix A), and the highest source in the 100-car naturalistic driving study (Neale et al., 2005).

Mobile phones have become essential equipment in peoples' day-to-day lives and they can be considered a ubiquitous technology with adult mobile phone ownership greater than 75%, and 90% for the 15 to 34 age group (National Statistics, 2006). In 2005, in the UK, 13 deaths and 400 injuries were attributed to the use of mobile phones whilst driving (BBC News, 2007).

Mobile phones are not designed specifically for in-vehicle use, yet drivers on the move regularly use them. Many countries, including the UK, have passed legislation aimed at reducing mobile phone related vehicle accidents, banning the use of hand-held devices (Young et al., 2003, RoSPA, 2002, Goodman et al., 1997). However, such legislation may lead to a belief amongst the driver

population that the use of hands-free equipment removes any associated risks of using a phone whilst driving. This is not the case. Several research studies have demonstrated that hands-free devices are no safer for concurrent use whilst driving than their hand-held counterparts (Young et al., 2003, Haigney et al., 2000, Redelmeier and Tibshirani, 1997).

There may be strong motivations for drivers to interact with their mobile phones whilst driving, with the use of such devices not as far removed from the overall driving task as initially thought. For example, if caught in traffic a driver may feel it necessary to telephone a family member/colleague etc. to inform them they are running behind schedule. Furthermore, local radio stations commonly call for drivers to become 'hands-free phone-rangers', calling in with information on traffic problems.³ The development of advanced mobile phones with Personal Digital Assistant (PDA) functionalities has seen devices that combine telephony with GPS navigation software. Similarly, the latest in-car navigation systems come equipped with Bluetooth technology, enabling drivers to 'pair' their system with their mobile phone.⁴ The navigation system may then be used as a hands-free car kit. Such functionality forms a significant part of the marketing for such devices. This all serves to blur the boundaries between in-car technology and devices not suited, or originally designed, for in-car use. Drivers receive mixed, conflicting, messages. Whilst governmental advice suggests drivers should

³ For example, Leicester Sound radio station. See <http://www.leicestersound.co.uk/travel>, accessed 10th August 2006

⁴ For example, TomTom GO 910, <http://www.tomtom.com/products/product.php?ID=212&Language=1>, accessed 10th August 2006.

avoid making phone-calls, and could still be prosecuted whilst using hands-free equipment for not having full control of their vehicles (DfT, 2006), manufacturers may be seen as condoning mobile phone use in the vehicle.

A growing body of literature has examined the effects of mobile phone use whilst driving (Goodman et al., 1997, Redelmeier and Tibshirani, 1997, Burch, 2002, Lam, 2002, RoSPA, 2002, Young et al., 2003). Whilst the general consensus appears to show an increased crash risk associated with mobile phone use (Lam, 2002), it can be difficult to draw firm conclusions (Young et al., 2003). Redelmeier and Tibshirani (1997) found that use of a mobile phone whilst driving increases the risk of having a collision by four times, similar to driving with a blood alcohol level at the legal limit. However, in their examination of accident statistics, Stutts et al. (2001) discovered no increase in the number of accidents involving mobile phones. It is suggested this may be related to the associated legal and financial issues of drivers admitting to using a mobile phone at the time of a crash.

The use of a mobile phone whilst driving may encompass aspects of each of the four distraction types (Young et al., 2003). They are a pronounced source of auditory distraction, greater than listening to a radio or conversing with a passenger (Young et al., 2003). In contrast to conversing with a passenger, mobile-phone conversations may suffer from sound quality variability (Young et al., 2003, Burch, 2002). More importantly, there is a lack of situational awareness and adaptation. Passengers allow conversations to lapse during demanding driving periods, as they are aware of what is going on in the road

environment. A person on the end of a phone-line is not aware of changing driving demands, and will continue to carry out a, potentially stressful, conversation (Young et al., 2003, Burch, 2002). Furthermore, passengers can become active participants in the driving task, through shared situational awareness and collaborative problem solving (Strayer, 2005). Hand-held phones are particularly physically distracting, requiring one-handed driving (RoSPA, 2002, Goodman et al., 1997), although the use of hands-free devices is not without associated physical demands (Young et al., 2003).

That there appears to be no discernable difference in crash risk between using a hand-held or hands-free mobile phone (e.g. Redelmeier and Tibshirani, 1997, Klauer et al., 2006), suggests that it may be the cognitive and auditory demands of carrying out a mobile phone conversation that are most disruptive to driving performance (Young et al., 2003). Alternatively, this may be attributed to the increased exposure to distraction when conducting a conversation (Klauer et al., 2006). Of potential concern, there is evidence that drivers are unaware of the risks they pose to themselves and others by using a mobile phone whilst driving (Lesch and Hancock, 2004).

2.3.6 Types of distraction

As stated, establishing distraction as the cause of an accident can be problematic (Stevens and Minton, 2001). Consequently, only in a minority of cases can distraction type (i.e. visual, cognitive, auditory, and biomechanical) be confidently established. Therefore, there is a scarcity of research establishing individual types of distraction to accident risk. However, recent studies suggest that visual distraction is a particular cause for concern. The

results of the 100-car naturalistic driving study demonstrated that 93% of all lead vehicle crashes involved inattention to the forward roadway as a contributory factor (Neale et al., 2005). Furthermore, Angell et al. (2006) found that the effects of cognitive distraction are subtle in comparison to visual distraction. Visual distraction was found to account for greater variance in driving performance, and visual-manual tasks were found to have a more pronounced effect on driving performance than audio-vocal tasks (Angell et al., 2006). Consequently, a focus on reducing visual distraction is recommended.

2.4 Summary

This chapter has presented an overview of the literature on driver distraction and its role in accident causation. In particular, it has focused on how driver distraction should be defined, and, the role of technology in distraction related accidents. It was suggested that a comprehensive definition of driver distraction has to consider four aspects: the distinction between distraction and inattention; that distraction sources can be internal or external to the vehicle; the four types of distraction; and, the effect of distraction on fundamental driving tasks. A proposed definition of driver distraction is:

Delay by the driver in the recognition of information necessary to safely maintain the lateral and longitudinal control of the vehicle (the driving task) due to some event, activity, object or person, within or outside the vehicle, that compels or tends to induce the driver's shifting attention away from fundamental driving tasks by compromising the driver's auditory, biomechanical, cognitive or visual faculties, or combinations thereof.

An overview of the role of distraction in accident causation was presented. Efforts to make comparisons between studies are hindered due to differing interpretations of driver distraction and disparity between data sources. Whilst

estimates suggest that between 25% and 50% of accidents involve some form of driver inattention, with distraction contributing to half of these (Stutts et al., 2001), key studies tend to produce much more conservative estimates (Wang et al., 1996, Stutts et al., 2001, Stevens and Minton, 2001). Contrastingly, a unique study of naturalistic driving behaviour found inattention contributed to 78% of accidents, with secondary task engagement the greatest cause of inattention (Neale et al., 2005). There is little doubt that distraction is a risk factor that drivers are exposed to, and there is scope for manufacturers of vehicles and in-car systems to minimise the potential for distraction through good design (Young et al., 2003).

Due to the apparent scarcity of firm data regarding distraction and accident rates, especially when considering more novel in-car systems, a significant branch of research has focused on the impact of distracted driving on performance measures. However, studies that focus on measurements of driving performance must be handled with care. It is important to establish the link between driving performance and increased crash risk (Simpson, 2005). Whilst performing a particular activity may be distracting *per se*, in reality drivers may employ coping strategies to minimise the increased crash risk (Young et al., 2003). Nevertheless, numerous studies have demonstrated that current and future in-car technology developments are a potentially worrying source of driver distraction (Goodman et al., 1997, Redelmeier and Tibshirani, 1997, Burch, 2002, Lam, 2002, RoSPA, 2002, Young et al., 2003, Klauer et al., 2006).

3 Literature review – Measuring distraction: evaluation methods

Chapter 2 addressed the issue of driver distraction and its involvement as a causal factor in road accidents. It was established that driver distraction plays a significant role in accident causation. Whilst there is a scarcity of epidemiological data regarding the role of in-vehicle technology, and, in particular, new forms of technology, e.g. IVIS, in accident causation, there is evidence that the further introduction of such technology may be a cause for concern. For example, studies have shown that for drivers using a mobile phone crash risk increases by as much as four times, whilst engaging in a complex secondary task can increase risk by three times (Redelmeier and Tibshirani, 1997, Klauer et al., 2006).

This chapter reviews the literature regarding research methods for assessing driver distraction. The review is broadly split into two sections: user trial methods and non-user trial methods. It is in the former category where the vast majority of research has focussed. However, as is shown, there have been some interesting efforts to develop non-user trial methods. Whilst a variety of methods are considered in the course of this review, particular attention is paid to the occlusion technique and the Keystroke Level Model (KLM), since these methods are most relevant to the experimental work of the later chapters.

Further, methods generally focus on the effects of visual distraction in particular. It could be argued that this type of distraction is the most concerning; intuitively, systems that require drivers to divert their visual attention from the forward road view are not desirable. Current IVIS predominantly rely on interaction through graphical user interface (GUI) menus and icons, in a similar fashion to modern desktop computing applications (Burnett et al., 2004). Such applications were designed for single-task environments where there are unlikely to be significant safety concerns; people rarely die or seriously injure themselves in the normal use of personal computers, for instance (Green, 2003). Consequently, such interface design paradigms may not be suitable for IVIS. The focus on visual distraction is not intended to imply that other types are not relevant or cause for concern, since evidence presented in the previous chapter suggests this is not the case. However, given that IVIS interfaces are currently reliant on visual-manual interaction, and that visual distraction appears to be of greatest safety concern (Klauer et al., 2006), the main focus of this research will remain visual distraction.

3.1 User trial methods

Thus far, distraction research has concentrated on the development of user trial based methods. Within HCI there is a well established tradition of using prototypes of varying degrees of fidelity throughout a design process, ranging from paper based to near fully functional systems (Preece et al., 2002). However, the safety-critical characteristics of the driver distraction problem demand that methods focus on measurements of human performance in IVIS

tasks. Such methods generally require robust, high-fidelity prototypes, which can cope with the rigours of testing, and therefore can be costly to run (Green, 2003, Lansdown et al., 2004, Salvucci et al., 2005). Thus, it is suggested that they may only be implemented in the later stages of development, when significant changes to the design may no longer be practical (Green, 2003, Salvucci, 2005). Consequently, the result of an assessment is likely to be a decision on whether a task is suitable for concurrent performance whilst driving or not, allowing little opportunity to redesign for such use. In effect, methods are more likely to be used for summative evaluation of designs, rather than to inform formative design. This situation is inconsistent with the goals of a user-centred iterative design process.

Trials may focus on measures of driver performance, and the effect of IVIS task performance on these measures. Variables include lateral position, speed maintenance and control, reaction times, glance behaviour, gap acceptance, and, workload. All have been measured in dual task studies to provide information on driving performance whilst attempting to perform secondary tasks (Young et al., 2003). Studies have used on-road and test track environments, and driving simulators of varying degrees of fidelity (Young et al, 2003). Alternatively, there are a few single-task methods where the focus is on the performance of IVIS tasks under specific conditions (Angell et al., 2006). Methods such as the 15-second rule and the occlusion technique, do not involve any direct assessment of driving behaviour or performance, for example. Such methods have drawn on previous studies of driving performance, and have been validated against established performance

measures, in order to establish criteria on which an in-vehicle interface can be judged (Green, 1999a, Young et al., 2003, Stevens et al., 2004, ISO, 2007).

In general, methods that focus on IVIS task performance have been more clearly defined than methods concerned with the effects of engaging in IVIS tasks on driving performance. Standards have been developed, or are in the process of development, for the implementation of the 15-second rule, the occlusion technique, and the lane-change test (LCT), and these techniques are generally well understood within the research community. For driving performance-focussed methods, there is greater variability. On-road and driving simulator studies have been described as methods (Young et al., 2003). However, the variance in experimental design, equipment and setup, and the potential measures obtained from such studies, means there is no guarantee that results from two simulator/on-road studies are comparable in the way that, for example, occlusion studies, generally are (assuming trials follow the ISO standard).

3.1.1 The 15-second rule

The 15-second rule was originally developed specifically for assessing navigation system tasks; however, it has since been used to assess a range of other IVIS tasks (Young et al., 2003). The aim of the rule is to simplify evaluations by making decisions based on one criterion, the total time to complete a task when in a stationary vehicle (static task time) (Young et al., 2003). Put simply, the rule states that functions with a static task time of 15 seconds or more should not be accessible to a driver whilst the vehicle is in

motion (Farber et al., 2000, Young et al., 2003).⁵ The rule has been standardised by the Society of Automotive Engineers (SAE), and is a recommended practice (SAE, 2000).

The 15-second rule is attractive to industry and standards organisations due to its simple implementation and procedure. However, it has attracted a certain degree of criticism (Tijerina et al., 2000, Young et al., 2003, Green, 2003, Salvucci et al., 2005). Research has suggested that, whilst the rule does identify the most distracting tasks, the 15-second value is relatively arbitrary – a 30-second rule would be just as effective, for example (Tijerina et al., 2000, Young et al., 2003). Particular criticisms of the rule focus on its inability to distinguish between tasks based on their level of visual demand (Stevens et al., 2004). A task that takes nine seconds to complete whilst stationary, but that requires the users' sustained visual attention throughout, would be deemed acceptable under the 15-second rule, although it would clearly not be suitable for drivers to engage in whilst on the move. The rule has been strongly defended by its authors (Green, 2003), and many researchers believe it fulfils its 'fundamental purpose of reducing the performance of tasks with long completion times whilst the vehicle is in motion' (Young et al., 2003: p. 37, Farber et al., 2000).

It is unlikely that the 15-second rule will deem tasks acceptable that are, in reality, less desirable. A task that takes 15-seconds or less is, in practice, not likely to require individual glances greater than two seconds duration (Green,

⁵ Participants performance in three instances of a task is averaged, and the mean of these is taken as the final result

2003). However, the rule may ‘fail’ tasks that are no more distracting than ‘acceptable’ tasks, despite their greater static task time. A well-designed interface may allow users to complete a task in brief chunks. Indeed, Chiang, Brooks and Weir (2001) observed no statistical differences in the number of lane deviations made when destination entry tasks were performed whilst driving, in comparison with driving only. This was attributed to the ability of drivers to complete the task in manageable chunks of interactions (Chiang et al., 2001). As such, despite benefiting from simplicity, it may be argued that the 15-second rule lacks the appropriate sensitivity necessary to assess potential distraction.

3.1.2 The occlusion technique

Visual occlusion has figured in driving research for a number of years. Originally employed to determine the workload of the primary driving task (Senders et al., 1966), the occlusion technique has more recently been investigated as a potentially useful measure of the visual demand associated with IVIS. Its premise lies in the assumption that, by allowing participants only brief periods of vision, the glancing behaviour of performing a secondary task in the driving environment is simulated. When engaging in a secondary task whilst driving, visual attention is shared between the primary driving task and the secondary, in-vehicle task. In occlusion trials, periods where participants are without vision may be viewed as representative of glances to the road scene. A shutter or goggles are used to simulate this behaviour by blocking and revealing the visual scene (Young et al., 2003, Stevens et al., 2004). If a task can be carried out effectively through a series of short glances,

it can be considered more acceptable to perform whilst driving, as this is an indicator that the task is highly 'chunkable', and relatively easy to resume (Stevens et al., 2004).

Interest in the occlusion technique has been driven by its low cost and ease of implementation in comparison to more traditional measures of driver distraction, such as road and simulator studies (Young et al., 2003, Stevens et al., 2004). Such studies require detailed planning (e.g. development/selection of appropriate road scenarios), time-consuming analysis (e.g. frame-by-frame video analysis), and expensive equipment (e.g. driving simulators/instrumented vehicles), whereas the occlusion technique can be implemented in a laboratory setting, or bench prototype phase (Stevens et al., 2004, Angell et al., 2006). In addition, the technique can be used to evaluate several system properties, such as 'chunkability, completion time, ease of resumption after interruption and visual complexity' (Young et al., 2003, p. 36). This diagnostic ability is highly desirable. In contrast to the 15-second rule, it is felt occlusion has the ability to discern between tasks that can be effectively completed in a series of short glances, and those that require long individual glances (Young et al., 2003, Stevens et al., 2004).

Common sense suggests that long glances away from the forward road view are undesirable. To support this, Klauer et al. (2006) demonstrate that single glances greater than two seconds translate to a significant increase in crash risk. However, the use of average single glance duration as a safety measure is problematic. Research has demonstrated that drivers tend to self-limit glance

durations away from the road scene to approximately 1.5 seconds (Rockwell, 1988, Wierwille, 1993). Whilst complex displays may encourage a tendency to longer single glances, it is suggested that such interfaces generally lead to higher glance frequencies and total eyes-off-road times than a significant increase in mean single glance duration (Wierwille, 1993). Thus, the occlusion technique seeks to filter out visually complex interface designs.

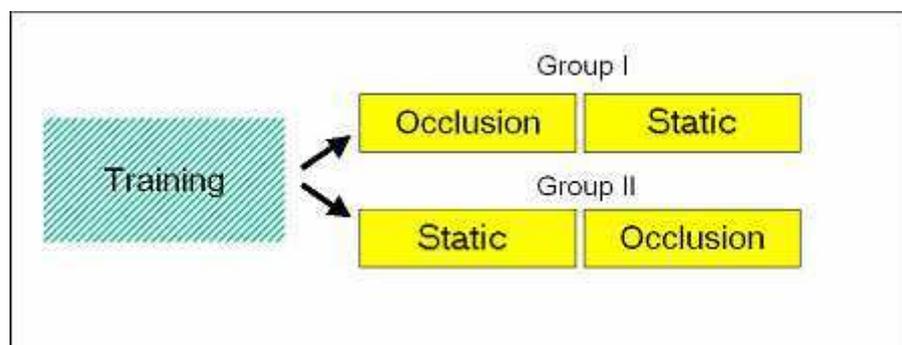


Figure 3.1 Experimental set up (ISO, 2007)

An international standard for the application of the technique has recently been published (ISO, 2007). Whilst there is potential for various configurations of the technique, and a number of different measures can be made and reported, the standard provides a framework for research and the application of the method (e.g. experimental design, training, reported measures), in order to improve cross-study comparisons.

In the standard, a number of conditions for study set-up are given. The key points are as follows:

- A minimum of 10 participants should be used in any occlusion study, with at least 20% of participants over the age of 50;
- Participants should undergo training in both IVIS tasks and the occlusion method;
- Up to 5 practice trials should be undertaken prior to actual trials. If participants do not reach an acceptable level of performance following practice trials, the training procedure and interface design should be reviewed;
- Participants should complete five trials on each task with full vision (static trials) and five trials whilst under occlusion conditions. The order of these conditions should be counterbalanced between participants in order to avoid learning effects (see Figure 3.1);
- Tasks acceptable for occlusion studies should have static task times of greater than five seconds;
- In occlusion conditions, periods where participants are allowed vision are referred to as shutter open time. Periods where participants are without vision are referred to as shutter closed times (see Fig. 3.2a and Fig. 3.2b).



Figure 3.2 Occlusion goggles: (a) shutters open; (b) shutters closed

The appropriate length of time participants are allowed vision and the duration of occluded periods has been the subject of much research (Stevens et al., 2004). Studies have demonstrated that shutter open periods of between 1.5 and 2 seconds are appropriate for assessing IVIS visual demand, with the Alliance of Automobile Manufacturers (AAM), ISO and Japanese standards organisations converging on a value of 1.5 seconds (AAM, 2003, ISO, 2007, Stevens et al., 2004). This value has been found to correlate well with actual glance times in on-road evaluations, and is felt to reflect a tolerable time for drivers to glance away from the road whilst in motion (Wierwille, 1993, Weir et al., 2003, Rockwell, 1988, Niiya, 2000). The duration of occluded intervals has similarly been investigated, with a broad range of times felt to be acceptable (Stevens et al., 2004). Early drafts of the standard proposed a shutter-closed time of 2 seconds and a shutter open time of 1.5 seconds (ISO, 2004a, ISO, 2005), however this has recently been revised, and the full standard proposes 1.5 seconds for both shutter-open and shutter-closed intervals (ISO, 2007). This is not considered a critical change, mostly implemented for the sake of simplicity, i.e. same occlusion/vision interval lengths.⁶

Occlusion metrics

The standard proposes that two metrics are reported following occlusion trials – the total shutter open time (TSOT), which is the total time that participants had vision whilst performing the task, and a further measure known as the resumability ratio (R) (ISO, 2007). The term resumability is something of a

⁶ Stevens, personal communication

compromise, with the metric originally being described as a measure of interruptability. However, since the measure seeks to indicate the ease with which participants are able to resume tasks following interruptions, rather than how easily tasks are interrupted, resumability is viewed as a more appropriate description. Furthermore, the measure might be further qualified by the admission that it is a reference to *visual* resumption of a task since the single-task nature of occlusion trials mean participants do not truly disengage from the task. This last point has not been commented on in the occlusion literature or the standard. In the full standard, R is an optional metric, and should only be used in conjunction with TSOT (ISO, 2007). Guidelines are provided for the calculation of these two metrics from the trials undertaken by each participant, and a rationale for the elimination of outliers is included. Examples of outliers are instances where a participant refuses to complete a trial, where the experimenter judges that a participant cannot successfully complete a trial, and where the total occluded task time exceeds the participant's average total static task time by a factor greater than four (ISO, 2007).

The TSOT metric is calculated by counting the number of shutter open periods and multiplying by 1.5, or approximated from the total occluded task time as follows:⁷

$$\text{TSOT} = (\text{total task time}_{\text{occluded}} / 3) * 1.5$$

⁷ In earlier drafts of the standard, where the shutter closed period was set at 2 seconds, the value in the given equation of 3 was, instead, 3.5

That this calculation is an approximation should be stressed, however. For example, a task with a total occluded task time of 4.3 seconds would, in reality, have a TSOT value of 2.3 seconds. The calculation approximation, following the above method, would yield a TSOT of 2.15 seconds.

Whilst it is necessary to measure total static task time for use in calculating R, the standard does not require that this measure be reported. Where R is not to be reported, static task time need not be measured. R is determined as the ratio of TSOT to total static task time (TTT_{static}), such that. :

$$R = (TSOT) / (TTT_{static})$$

The values used in the calculation of R are taken from the mean TSOT and TTT_{static} values obtained from the five task instances completed in each condition, yielding a single R-value for each participant.

Validity

A number of studies have sought to assess the validity of the occlusion technique (Niiya, 2000, Hashimoto and Atsumi, 2001, Young et al., 2003, Stevens et al., 2004, Baumann et al., 2004). Whilst individual research has demonstrated that the technique shows promise, there is concern that there has been a lack of coordination between studies (Young et al., 2003). Correlations have been found between occlusion measures and a number of driving performance measures, implying that occlusion trials may be a valid substitute for on-road and simulator studies (Stevens et al., 2004).

A number of different measures have been reported in occlusion studies, including: TSOT, R, total task time in static and occluded conditions, and a ratio, $TTT_{occluded}/(TTT_{static}+T_{tclose})$, which compares static total task time (TTT) with occluded total task time ($TTT_{occluded}$), taking into account the amount of time participants were without vision (T_{tclose}) (Stevens et al., 2004). Measures of total task time have been shown to correlate well with TSOT, whilst the ratio $TTT_{occluded}/(TTT_{static}+T_{tclose})$ is described as being similar to R (Stevens et al., 2004). The ISO standard defines only the TSOT and the R ratio parameters (ISO, 2007). Consequently, discussion will focus on these metrics.

Total Shutter Open Time (TSOT)

Studies have shown that TSOT correlates highly with total glance time – the total time drivers spend glancing away from the road whilst driving (Niiya, 2000, Hashimoto and Atsumi, 2001, Bengler, 2006, Shutko and Tijerina, 2006). As TSOT is much more easily obtained than total glance time, this is encouraging (Stevens et al., 2004). TSOT has also been shown to be repeatable, statistically discriminating, and predictive of on-road task time, standard deviation of lane position (SDLP), total glance time to an IVIS, and the number of glances to an IVIS (Angell et al., 2006, Tijerina, 2006b). Recommendations regarding acceptable values for TSOT have been made; however, research suggests that many IVIS tasks seriously exceed these (Stevens et al., 2004). Total glance time has also been shown to correlate well with static task time with full vision (Green, 1999a); however, studies have

suggested correlations are stronger between TSOT and total glance time (Angell et al., 2006, Asoh and Iihoshi, 2006).

Total glance time has been considered a reliable driving safety measure, with research showing strong correlations between total glance time and lane deviations, also a driving safety measure (Niiya, 2000, Curry et al., 2002, Shutko and Tijerina, 2006). However, it is argued that there are problems with assessing IVIS based on total glance time; high values imply the IVIS is unsafe, yet if a task is highly chunkable, and requires only short glances away from the road whilst driving, it may be possible to accomplish it with minimal impact on the primary driving task (Stevens et al., 2004). This is perhaps a problematic viewpoint, since it implies that tasks of any length are acceptable if they can be effectively chunked. Therefore, it presumes it is 'safe' for drivers to read a novel whilst driving, so long as they do so six words at a time.

⁸ This does not account for the increased exposure to risk, which leads to a greater likelihood of being involved in a crash (Klauer et al., 2006).

As has been noted, it is suggested that measures of static task time alone, such as those obtained with the 15-second rule protocol, lack sensitivity, as they fail to discriminate between tasks that require long glances and those that can be completed with only short glances (Young et al., 2003, Stevens et al., 2004). Similar issues have been raised regarding total glance time. In an on-road study, Chiang et al. (2001) found that participants made a similar number of lane exceedences when performing a destination entry task whilst driving as in

⁸ Burns, P., personal communication

a baseline condition (driving only, with no secondary task to perform). This was attributed to the highly interruptible nature of the secondary task. Thus, a task with a long total task time, or total glance time value, may be acceptable if individual glances are short. Logically, a task with a total glance time of 18 seconds comprised of 18 individual, one second glances, is more desirable than a task with a total glance time of nine seconds comprised of, for example, three long glances. As TSOT correlates well with total glance time, it can be inferred that the measures may have similar failings. Therefore, it is proposed that static task time, TSOT and total glance time are not necessarily indicative, in isolation, of the overall impact of performing a secondary task on driving performance (Stevens et al., 2004). However, TSOT improves on the 15-second rule by measuring, essentially, the total visual time required to complete a task. Therefore, a task with a long static task time that affords significant non-visual interaction will perform more favourably in a TSOT assessment than a shorter task that requires complete, focussed visual attention, or does not afford opportunities for blind control.

Resumability ratio (R)

Misgivings regarding the completeness of the TSOT measure led to the development of the resumability ratio (R), which is used to evaluate ‘the degree to which a task can be done without visual control’ (Stevens et al., 2004: p. 25). Care should be taken with the measure as it can be negatively affected by the static task time of an IVIS task. The ISO standard suggests that the occlusion technique is only suitable for tasks longer than five seconds duration, since R has been found to be influenced more by static task time

when static task time is short, than by task complexity (Goujon, 2001, Karlsson and Fichtenberg, 2001).

Resumability is a complex characteristic, influenced by several factors (Weiner, 1972, Tijerina, 2006a):

- **Activity type.** The nature of the task will have an impact on the tendency to resume. For example, tasks with finite goals are more likely to be resumed than tasks with indefinite ends.
- **Task progress,** or, place in the task at the point of interruption. As a person gets closer to the end goal, they are more likely to resume the task following an interruption.
- **Length of interruption.** There is a greater likelihood of resumption following short interruptions as opposed to long interruptions.
- **Individual attitudes.** Where a person is intrinsically motivated to perform a task there is a greater likelihood they will resume it following an interruption than if they have been induced.

In an occlusion assessment, many of these factors are addressed. Tasks chosen for assessment are required by the draft standard to have a definite end goal and interruptions are set at fixed intervals. Finally, whilst participants may not be intrinsically motivated to perform the task as such, they are motivated by outside factors, evidenced by their willingness to take part in the evaluation.

In an assessment of the effectiveness of the R metric, Tijerina (2006a) found that R did not correlate with the measures of glance frequency, total glance time, or mean single glance duration. Furthermore, R was not found to

correlate with measures of driving performance, such as typical speed difference and Standard Deviation of Lane Position (SDLP). However, the metric was found to correlate with some object and event detection measures. Significantly, TSOT was not found to correlate well with these latter measures. Similarly, Bengler (2006) also demonstrated a relationship between R and measures of object and event detection; however, results are not reported in depth. Bengler (2006) found that R correlates well with lane keeping behaviour and further concluded that the metric is able to differentiate between IVIS based on their display complexity, including different levels of dialogue dynamics, levels of interaction timing and temporal behaviour and between IVIS tasks based on the ease of resumption following an interaction. Whilst confidence in the metric is not particularly high, the recent research of Bengler (2006) and Tijerina (2006a) suggests that R may be of some use; it may be that the diagnostic ability of R is useful in identifying opportunities for reducing TSOT (Burns et al., 2006b). More recent drafts of the standard, and the final published standard, have seen the R metric relegated to an optional measure only, specifically, R can be used in conjunction with TSOT but never alone (ISO, 2006, ISO, 2007, Burns et al., 2006b).

Acceptable values for TSOT and R

The ISO standard does not indicate an acceptable value, or limit, for TSOT (ISO, 2007). However, limits have been defined by other organisations. The AAM suggests that tasks where TSOT is less than 15 seconds are acceptable (AAM, 2003). Their research suggests that this value corresponds with a total glance time when driving of less than 20 seconds, which is felt to be tolerable.

Alternatively, Asoh and Iihoshi (2006) claim that a total glance time value of eight seconds is acceptable based on correlations with drivers feelings of uneasiness whilst performing a secondary task and vehicle's lateral deviation. The corresponding TSOT value for a total glance time of eight seconds is 7.1 seconds. The Driver Workload Metrics project drew similar conclusions, with research suggesting that mean TSOT values greater than 7.5 seconds correspond to visual-manual tasks of high workload (Angell et al., 2006).

For R, the ISO standard suggests that values greater than one may indicate that participants are having difficulty in resuming a task following interruptions (ISO, 2007). The standard does not formally state, however, that tasks where R exceeds this value should be classed unacceptable. Few studies have presented results where mean R-values exceed one. In the driver workload metrics project, where thirteen visual-manual tasks were examined, just one produced a mean $R > 1$ (Angell et al., 2006). The characteristics of this task, hard map reading, correspond to the notion of a task that drivers might have difficulty interrupting and resuming. The authors suggest drivers may have to invest extra visual effort in order to reorient themselves to the task following interruptions (Angell et al., 2006). Similarly, Noy et al. (2004) report mean R-values greater than one for two tasks that required a significant amount of visual searching of a display. A tentative conclusion might be to only consider the R metric relevant where its mean value exceeds one.

3.1.3 Lane change test (LCT)

The lane change test (LCT) has been developed by Daimler-Chrysler as a low cost method for assessing driving performance whilst engaging in secondary

tasks, and is currently in the early stages of standardisation (Mattes, 2003, Mattes, 2006). Participants are required to drive a low-fidelity, desktop-based simulator whilst performing tasks on an in-vehicle device (see Figure 3.3). Signs on the sides of the simulated roadway instruct drivers that they should change lanes (see Figure 3.4).

Participants' lane-change performance is assessed against a normative model (see Figure 3.5). In the figure, the green (angular) line indicates the normative trajectory, the blue (curvy) line is the driver's actual trajectory and the red (shaded region) indicates the area in between the two.

As such, the method is sensitive to performance based on: perception, reaction, manoeuvring and lane-keeping (Burns et al., 2006b). In this way, the method has the ability to assess a driver's awareness of the driving environment (perception, reaction), and, their ability to safely control the vehicle (manoeuvring, lane keeping). Poor performance in any one of these areas, or combinations thereof, leads to greater deviations from the normative model, therefore increasing the red area (Mattes, 2003, Burns et al., 2006b). As such, the LCT could be seen as an effective measure of both visual and cognitive loading. Visual tasks have been described as mostly leading to reduced lateral control, whilst cognitive tasks mostly lead to late, or erroneous, detection, both of which would be detected by the LCT (Engström and Markkula, 2006). An adapted version of the test, measuring average distance, lane change length and start of lane change in comparison to a normative model has been shown to

provide more accurate results than the conventional assessment measures (Veste and Bruyas, 2006).



Figure 3.3 Lane-change task set-up (from Burns, 2006, used with permission)

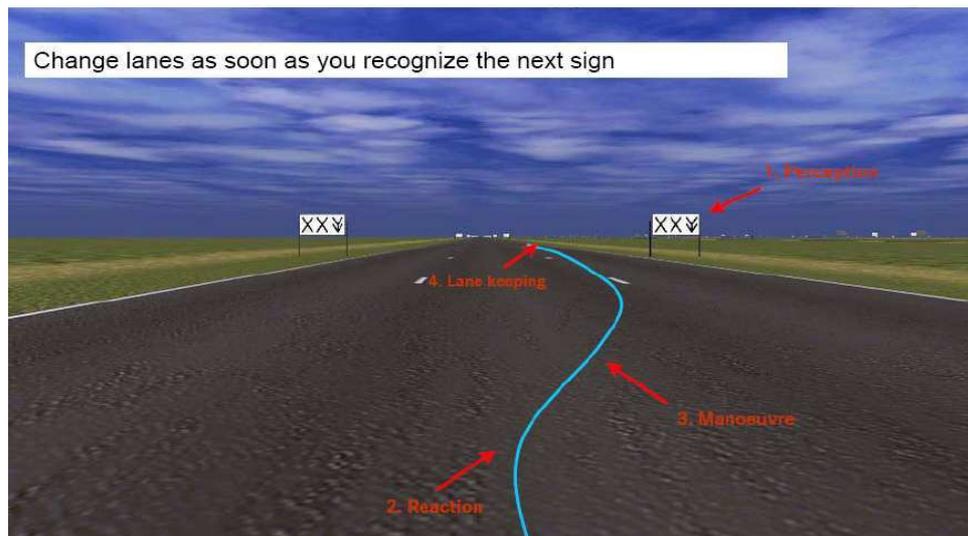


Figure 3.4 Example screen in LCT (Mattes, 2003, used with permission)

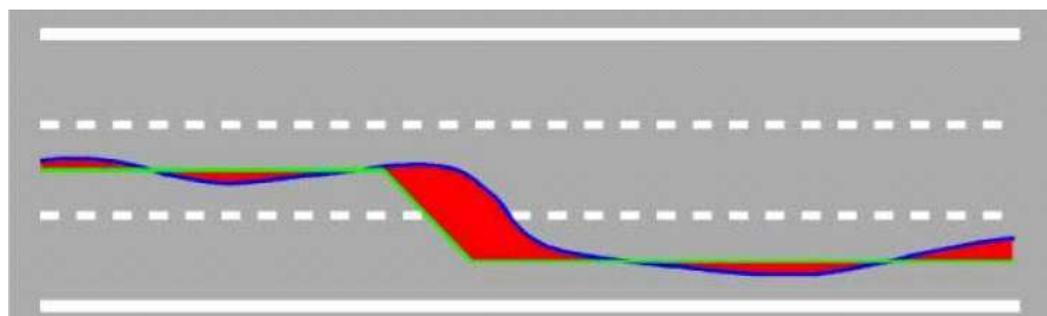


Figure 3.5 Example of LCT performance (Mattes, 2003, used with permission)

The LCT has the potential to be a very useful method. It allows a simple measure to be taken from simulator studies, enabling easy cross-study comparison. The method is also practical, suitable for standardisation, with good face validity and the potential to assess several driving performance factors (Burns et al., 2006b). However, there are a few potential limitations to the LCT. There is some degree of uncertainty regarding the method, the metrics and the criteria (Burns et al., 2006b). The sensitivity of the primary measure, the average deviation from the normative model, is unclear, and, furthermore, there is a lack of published research on the method (Burns et al., 2006b).

The focus on measurements of task performance means that, similar to other user-trial methods, the LCT requires well-developed prototype systems. Furthermore, acceptable limits need to be established if the method is to have reasonable utility. Whilst research has evaluated performance in the LCT whilst engaging with secondary tasks in comparison to performance when driving only (Burns et al., 2006b), this could be placing unreasonable expectations on drivers and/or system performance. It is unlikely that an interface, however well designed, will have *no* detectible effect on driving performance. What is important is to establish what level of effect is permissible.

A further issue with the LCT is the lack of published material evaluating its effectiveness. The method is in the early stages of standardisation, and some

promising results have been presented at workshops (e.g. Burns et al., 2006b, Engström and Markkula, 2006, Burns, 2006, Veste and Bruyas, 2006). However, there is currently a scarcity of peer-reviewed research in key journals.⁹

3.1.4 Peripheral detection task

Similar to the occlusion technique, the Peripheral Detection Task (PDT) was initially investigated as a method for determining the workload and attention demands of driving itself (van Winsum et al., 1999, Young et al., 2003). A PDT study requires participants to engage in a primary task (in this instance driving) whilst detecting and responding to targets (commonly lights), which are presented randomly in the peripheral areas of their visual field (Young et al., 2003). As the demands of the primary task increase, reaction time to peripheral targets increases. Participants may even fail to detect some targets if the primary task workload is high (Olsson and Burns, 2000, Young et al., 2003).

In general, PDT-based distraction studies have involved participants driving, in either a simulator or real vehicle, and interacting with IVIS whilst responding to PDT targets (Olsson and Burns, 2000, Harms and Patten, 2003, Young et al., 2003). Comparisons can then be drawn with target detection whilst driving only (Angell et al., 2006). Support for the method as a measure of visual distraction is strong, however its ability to distinguish between conditions due to variations in cognitive demand is less well established (Harms and Patten,

⁹ Mattes, personal communication

2003, Young et al., 2003). Furthermore, the lack of standardisation with the method is problematic for drawing comparisons between studies.

3.1.5 Simulator and on-road studies

A large body of driver distraction research has focussed on establishing the effects of distraction on driving performance. A variety of performance measures exist, and Table 3.1 presents a summary. The environment used to assess driving performance varies from low-fidelity driving simulators to on-road evaluations and field studies (Young et al., 2003). At the high-end of this range is the field study. Such studies involve equipping vehicles with instrumentation to monitor selected variables, in setups similar to the 100-car naturalistic driving study (Young et al., 2003, Neale et al., 2005, Klauer et al., 2006). Examples of field studies are extremely rare in this particular research area, although there are exceptions (e.g. Foley and Hudak, 1996). Whilst such studies can provide extremely rich data, there are associated problems. In addition to considerations of cost and logistics, it may be difficult to include a representative sample.

Short-duration on-road evaluation and test-track studies are more commonly performed (Green et al., 1993, Eby and Kostyniuk, 1999, Burnett and Porter, 2002, Cooper and Zheng, 2002, Harbluk et al., 2002, Young et al., 2003). These are more economical to perform than long-term studies, yet still yield results from real world driving (Young et al., 2003). However, there will necessarily be some level of artificiality, and this may have an effect on the naturalistic behaviour of participants.

Table 3.1 Driving performance variables and measures (adapted from Young et al, 2003).

Variable	Example Measures
Lateral Position	No. of lane deviations & exceedences, deviations from correct lane position
Speed Maintenance and Control	Variation in speed and throttle control, comparison of mean speeds
Reaction Times	Reaction time to an external event or object (e.g. traffic lights, braking lead vehicle)
Gap Acceptance	Comparison of gap acceptance in two conditions (whilst performing a secondary task and normal driving)
Workload	NASA TLX
Attention to Safe Driving Practices	Observation of time spent looking to mirrors and instruments

Due largely to the costs and risk factors involved in on-road studies, the most common environment for examining driver distraction is driving simulators (Young et al., 2003). These offer numerous benefits to researchers. They provide a relatively realistic, safe, and controlled environment in which to assess driving performance, and are generally more cost-effective than instrumented vehicles (Young et al., 2003). Driving simulators provide an ideal environment for research that may be unsafe to conduct on real roads, or in real vehicles (Goodman et al., 1997, Young et al., 2003). In the virtual environment, the possibility of crashing into another vehicle is a less important concern than in the real world. Furthermore, researchers can more precisely configure the driving environment with simulators, specifying the road layout, traffic and weather conditions, and so on (Reed and Green, 1999, Young et al., 2003). Correspondingly, changes to such test conditions can be implemented with relative ease (Young et al., 2003).

There are certain disadvantages of using driving simulators in distraction research. Whilst they may be more cost-effective than on-road evaluations,

high-fidelity simulators would rarely be described as cheap, especially in comparison with other equipment, such as the goggles used in occlusion studies (Reed and Green, 1999, Young et al., 2003). Other concerns include simulator sickness and learning effects (Goodman et al., 1997, Young et al., 2003). However, the most significant concern regarding driving simulators is the question of validity; to what extent does driving behaviour in a simulator reflect behaviour in a real vehicle? It is possible that participants will modify their natural behaviour because of being in the virtual environment. As there are less severe consequences to having a crash, drivers may view primary driving tasks as being of lower importance (Young et al., 2003). This has serious implications when participants are asked to engage in IVIS, and other, tasks whilst driving, since they may be less concerned with the effects of interacting with an IVIS on their driving performance than they would be when driving on real roads.

A number of studies have addressed the issue of simulator validity (Blaauw, 1982, Triggs, 1996, Reed and Green, 1999, Godley et al., 2001, Young et al., 2003). Two aspects of simulator validity have been proposed (Blaauw, 1982, Young et al., 2003):

- **Physical validity**, or the **fidelity** of the simulator, refers to the physical correspondence between the simulator vehicle and a real-world counterpart. This includes physical aspects, such as controls, seating, and so on, and, the virtual environment's level of realism. Whilst it may be possible to conduct some forms of research on low-fidelity simulators, for driver distraction research a reasonably high level of

fidelity is recommended, as the intention is to use the results to draw conclusions regarding driving performance in the real world (Triggs, 1996, Young et al., 2003).

- **Behavioural validity** refers to the extent to which driving performance in the simulator matches driving performance in a real vehicle. Two levels of behavioural validity are absolute validity and relative validity (Young et al., 2003). The former term applies when results from simulator studies are identical, or near enough identical, to results from an on-road evaluation, and is rarely demonstrated (Blaauw, 1982, Carsten et al., 1997, Reed and Green, 1999, Godley et al., 2001, Young et al., 2003). The latter term, relative validity, applies when results from a simulator study are similar to those of an on-road evaluation, although are not *exactly* the same. This level of validity is routinely demonstrated (Young et al., 2003).

There is significant variation between studies focussed on measures of driving performance. A multitude of different simulators, simulator set-ups, procedures, tasks, and performance measures have been used or assessed. This variability makes comparisons between studies difficult, and it is hard to nominate any one particular driving performance variable as being particularly susceptible to distraction. Instead, research suggests that many are affected (Young et al, 2003). In addition, certain devices or activities can have affects on multiple performance measures, creating a ‘cocktail for disaster’ (Young et al, 2003: p. 28). From the research that is available, however, it is suggested

that maintenance of lane position, speed maintenance and reaction times are particularly susceptible to the effects of driver distraction (Young et al, 2003).

Whilst research has focussed on dual-task studies comparing the effect on driving performance measures whilst performing a secondary task against driving with no secondary task, acceptable limits for driving performance have not been set. When considering the effects of a distraction source on lateral position and speed it has been argued that a certain amount of deviation is acceptable. Maintenance of lane position and speed does not have to be precise in order to constitute safe driving (Cnossen et al., 2004). A future research aim might be to establish baseline acceptable levels of driving performance against which new IVIS could be assessed. Those devices, or tasks, that are likely to impair driving beyond these limits would be deemed unacceptable in their present state for concurrent performance whilst driving.

Studies have focussed on the effect on driving performance in the presence of secondary tasks of varying degrees of difficulty. However, there is a lack of research examining the opposite – i.e. secondary task performance when the driving task varies in difficulty (Young et al, 2003). Cnossen et al. (2004) view driving as a hierarchical task, and suggest that some secondary tasks will fall higher into the hierarchy than others. They demonstrated that drivers are more willing to accept detriments to their driving performance when a secondary task is deemed as being of relevance to driving, e.g. route finding, but are not so concerned when the secondary task is unrelated to driving. Other work has shown that, in real world conditions, drivers seem to engage in

secondary tasks at what they perceive to be 'safer' times (Young et al, 2003: p. 21).

Assessing visual distraction in simulator and on-road evaluations

Visual distraction is commonly assessed in simulator and on-road evaluations through studies of participants' glance behaviour (Farber et al., 2000, Haigney and Westerman, 2001, Curry et al., 2002, Young et al., 2003). Traditionally, video cameras installed in the vehicle record the participants' eye movements throughout an experiment, and this is later analysed (Farber et al., 2000, Young et al., 2003). Recent developments in eye-tracking technology have reduced the need for frame-by-frame video analysis (Young et al., 2003). However, such systems do have associated costs: they are expensive, and may be difficult to install and calibrate (Farber et al., 2000, Young et al., 2003).

Typically, eye-glance studies record the duration and frequency of glances away from the roadway scene (Young et al., 2003). The resultant measures, total eyes-off-road-time, single glance duration, and glance frequency, combine to indicate the level of visual distraction experienced when performing a secondary task whilst driving. Total eyes-off-road-time has been found to correlate well with the number of lane excursions made whilst performing secondary tasks, a measure of driving performance (Curry et al., 2002, Haigney and Westerman, 2001, Young et al., 2003).

There have been efforts to determine acceptable levels of visual distraction. The AAM proposes that maximum single glance durations should not exceed

two seconds, and total glance time (eyes-off-road-time) should not exceed 20 seconds (AAM, 2003). These values are derived from research concerning the visual distraction associated with manually tuning a radio (AAM, 2003). Such a task has established safety relevance, being a distraction source listed in crash databases (see Wang et al., 1996, Stevens and Minton, 2001, Stutts et al., 2001, Young et al., 2003). Furthermore, such a task is said to be typical of in-vehicle device interaction (AAM, 2003). 85th percentile values are used to draw conclusions regarding maximum acceptable values (AAM, 2003).

A limit of two seconds maximum single glance duration follows research by Rockwell (1988). Studies of glancing behaviour using more than a hundred subjects over a six-year period demonstrated that drivers are reluctant to glance away from the forward view for much more than 1.5 seconds. A distribution of single glance durations based on this research, but incorporating more data (from a 10 year period), shows that the 85th percentile of glance durations in a histogram corresponds with a glance duration of approximately 1.9 seconds, which is rounded up to 2.0 seconds in order to provide a whole number engineering criterion (AAM, 2003).

The requirement is not considered sufficient in isolation. It has been suggested that safety-relevant measures, such as number of lane exceedences, are not well predicted by analysis of single glance durations (Green, 1999c). As noted earlier, this is likely due to self-regulation on the part of the driver (Rockwell, 1988, Wierwille, 1993, Green, 1999c, AAM, 2003). However, the recent findings from the 100-car naturalistic driving study challenge this position,

demonstrating that glances in excess of two seconds significantly increase crash risk (Klauer et al., 2006). Further, Rockwell (1988) suggests that for visually complex displays, drivers tend to increase glance frequency rather than duration. This is considered to have a significant impact on safety, therefore design should seek to minimise exposure to visually complex displays.

Several factors have been considered to determine the limit on total glance time of 20 seconds (AAM, 2003). Research has demonstrated a link between glance frequency and lane exceedences (Green, 1999c), which is not thought to be due to the increased task time associated with higher glance frequencies (AAM, 2003). However, a limit on glance frequency is considered insufficient, as many systems, particularly new technologies, may require a high number of very short glances to ascertain system status (AAM, 2003). Thus, a limit on total glance time is proposed. The limit itself is drawn from a study linking glance frequency, glance durations and lane exceedences (Dingus, 1987, Dingus et al., 1989, AAM, 2003). For radio tuning, the 85th percentile value for glance frequency is 9.4 (AAM, 2003). To reach the proposed limit of 20 seconds, this figure is rounded up to 10 glances and multiplied by the 2-second single glance limit (AAM, 2003).

Ethnographic studies of driving

Ethnography is a well-known technique within HCI research that has so far been under-utilised in driving and transport safety related research. A recent exception is a study of drivers using mobile phones by Esbjörnsson et al. (2007) in Sweden. The authors argue that the predominantly experimental

approach of current driver distraction evaluation methods is problematic since it does not consider the overall context of system use, and is therefore too narrow in scope to truly inform road safety policy. They claim this is evidenced by the discrepancy between the risk attributed to mobile phone use in experimental studies and the relatively low proportion of accidents caused by mobile phones in accident statistics. An ethnographic approach, whereby drivers' behaviour is observed in realistic settings (i.e. tasks instigated at the drivers' will, driving in real traffic situations) can inform understanding of driver behaviour and provide guidance for future system design. In their observations of drivers using mobile phones, Esbjörnsson et al. (2007) determine that drivers are likely to moderate their behaviour through a process they term *interactional adaptation*. Drivers were observed commenting on the complexity of the driving task whilst conversing on their phones, taking advantage of less demanding driving situations to instigate calls, and parking their vehicle when faced with extra demands, e.g. when they need to write something down.

Whilst it is perhaps fair to criticise current research for being biased towards experimental studies, the ethnographic approach comes with its own significant limitations. A particular concern with Esbjörnsson et al.'s (2007) study is that at no point do the authors comment on the driving performance of their subjects. In the narrative descriptions presented, for example, we see a driver regularly driving with neither hand on the wheel, due to the physical demands of holding his phone, for which no criticism is made. The inference appears to

be that since the driver concerned didn't have a crash he was therefore driving safely.

Esbjörnsson et al.'s (2007) motivation to understand driver behaviour in naturalistic settings is nonetheless worthwhile; however, the ethnographic methodology is perhaps not best suited. In contrast, it could be argued that the 100-car naturalistic driving study discussed in Chapter 2 follows a more useful approach (Neale et al., 2005, Klauer et al., 2006). Of particular relevance to the Esbjörnsson et al. (2007) are the findings relating to the proportion of accidents attributable to distraction, and the risk factors for engaging in secondary tasks. Esbjörnsson et al. (2007) argue that accident statistics regarding mobile phone related crashes and the experimentally determined risk factors do not match, and attribute this to the interactional adaptation of drivers that allows them to minimise their accident risk. However, the results of the 100-car study suggest that a much higher proportion (22%) of accidents is caused by secondary task engagement than accident statistics tend to show. An alternative hypothesis, therefore, is that distraction related accidents tend to occur at low speeds, causing minor damage, and are therefore not captured in police reports and other accident database sources.

In summary, ethnographic approaches to driver distraction may provide useful information about driver behaviour. However, they should always be designed such that observational findings are coupled with an examination of established measures of driver performance.

3.2 Non-user trial methods

The range of user-trial methods for evaluating driver distraction is impressive. Between them, these methods have the potential to consider a variety of driving performance effects of distraction. However, all these methods share one key limitation. Their assessment, through performance-based user trials, means that they can only, realistically, be implemented in later design stages, when at least a bench prototype is available (Angell et al., 2006). Consequently, their usefulness in the practical development of IVIS is questionable (Green, 2003, Salvucci et al., 2005). To this end, a driver-centred design approach must consider alternative methods that can be applied in early design stages. The following section details a number of non-user based methods, including qualitative and quantitative techniques.

3.2.1 Qualitative techniques

There is a wealth of design guidelines, standards and codes of practice available to IVIS manufacturers (Stevens et al., 2002). In the UK, the Department for Transport commissioned the Transport Research Laboratory (TRL) to develop a concise, practical and user friendly set of guidelines for the design of IVIS (Stevens et al., 2002). The procedure involved a comprehensive review of existing and emerging guidelines, and surveys of companies and individuals working within IVIS production and design (Stevens et al., 2002). Whilst the guidelines were developed for a specific scope – the focus is primarily on systems that provide drivers with journey specific information for instance – it is suggested that their use may be more general (Stevens et al., 2002). Furthermore, compliance with the guidelines

should result in a favourable evaluation by the TRL checklist (Stevens et al., 2002).

Good guidance is of immense value to IVIS designers. Tools such as the TRL checklist provide manufacturers and safety practitioners with an effective subjective evaluation technique to assess designs. However, their output is not quantitative in nature, and this has associated problems. Quantitative measures provide numerical results, which can be translated into engineering criteria (AAM, 2003). The core concern is how we *measure* distraction. Fundamentally, this question demands solutions that are quantitative in nature.

3.2.2 HCI and modelling behaviour

Providing quantitative information about user performance with interfaces relies on modelling techniques that anticipate or predict human behaviour. Within the HCI community there have emerged a few such methods, to varying degrees of success and acceptance. The following section discusses one such technique in detail – GOMS – which is arguably one of the most well known and established modelling paradigms. However, alternative prescriptive techniques exist and are worth mentioning, an example of which is using Markov modelling to evaluate the usability of devices in the early design stages (Thimbleby et al., 2001).

Markov models provide a means for mathematically modelling interactions with discrete devices, thus inferring measures of the difficulty users may encounter in achieving their goals (Thimbleby et al., 2001). Thus far, to the

author's knowledge, they have not been applied specifically to the driving domain. Further, they are generally well suited only to devices that have clear states and transitions, and are unable to consider continuous interaction (Thimbleby et al., 2001). In their study, Thimbleby et al. (2001) demonstrate how Markov modelling can be used to slight changes to interface design can affect its usability, how designers' knowledge can bias usability evaluations, and claim that the technique is applicable to a wide range of systems. However, a key limitation, recognised by the authors, is that Markov modelling is a mathematical technique which might not be well accepted by interface designers. The solution proposed is that the mathematics become automated within software tools enabling designers to input interface characteristics and extract usability measures without any 'craft knowledge' (Thimbleby et al., 2001, p. 123).

3.2.3 GOMS and KLM

GOMS models are the result of work by Card, Moran and Newell in the 1980s concerned with modelling human behaviour, with a view to providing quantitative predictions of behaviour that might be useful in system design (Card et al., 1983, John, 1995, John, 2003). A GOMS analysis breaks a task down in terms of the *Goals*, *Operators*, *Methods* and *Selection rules* needed to successfully complete the task.

- **Goals** are the users' overall anticipated outcome from using a system. For example, with an in-car navigation system, the goal may be to obtain route guidance to a particular location. For complex systems, goals may be more difficult to determine.

Interviews with potential users and observations of similar systems may be useful in these cases (John, 1995).

- **Operators** are the actions needed to achieve a goal as allowed by the system. These can be defined in a number of different ways, but are commonly at the ‘keystroke level’ and refer to button presses, cursor movements etc (John, 1995, John, 2003).
- **Methods** are the groupings of sub-goals and operators that in combination accomplish a goal (John, 1995, John, 2003).
- **Selection rules** are implemented whenever there is more than one method of accomplishing a goal. These identify the conditions under which particular methods will be chosen (John, 1995).

A GOMS analysis typically considers expert users, where task performance does not require problem solving (John, 1995, John, 2003, Raskin, 2000). The focus on skilled performance may appear limiting initially; however, a number of systems are designed with the intention that users will eventually become expert in their use. A GOMS analysis can assist designers in providing users with ‘efficient methods for accomplishing routine tasks’ (John, 1995, p. 58). A version of GOMS, Natural GOMS language (NGOMS), accounts for non-expert behaviour, i.e. learning times (Raskin, 2000).

GOMS techniques, and related methods, have been the subject of over one hundred published research papers (John et al., 2004a). However, despite demonstrations of the techniques’ validity and predictive value, modelling has failed to become a widely used tool by user interface designers (John, 2003).

This has been attributed to the effort required to learn how to apply the techniques, and the lack of suitable tools (John, 2003, John et al., 2004a).

Even without implementing a formal GOMS analysis, designers equipped with knowledge of the fundamentals behind GOMS ‘inherently incorporate GOMS teachings’ into their work (Raskin, 2000, p. 72). Whilst this may be true, formal analysis remains useful for choosing between designs, as the quantitative predictions made can highlight small differences that may relate to significant economic and psychological costs (Raskin, 2000, John, 2003). Research and implementation of GOMS models have demonstrated the technique’s ability to make highly accurate predictions of task performance times, adding weight to the technique’s validity (Raskin, 2000, John, 2003).

In a world where computing technology is becoming increasingly ubiquitous, there may be concerns that techniques such as GOMS, which was developed for desktop computing situations, are no longer valid. However, GOMS has been implemented across a number of different types of software and system, including the aviation field, with programmable television sets, and, active systems, including video games (John, 1995, John, 2003).

The most striking benefit of GOMS techniques is their ability to provide designers with quantitative information about proposed systems without the need to run costly user trials, which may come too late in a design process to be of substantial benefit (Dix et al., 2003, Green, 2003, Salvucci et al., 2005). The simplest variant of GOMS, the keystroke level model (KLM) is used, specifically, to predict the time taken by expert users to complete tasks on a

system (Raskin, 2000, John, 2003, Luo and John, 2005). KLM techniques involve decomposition of a task into its primitive operators, for example, pressing a key, pointing with a mouse, and so on. A time value is assigned to each operator, with the predicted task time being the sum of these values (Card et al, 1983, Raskin, 2000, John, 2003, Luo and John, 2005). A cognitive operator (M) is included to account for the time users take to mentally prepare for an action. A set of heuristics is provided to guide modellers in the placement of mental operators (Card et al., 1983, Raskin, 2000). It has been suggested that, when the KLM rules are applied rigorously and consistently, model predictions for skilled execution time should be well within 20% of observed results (John et al., 2004b).

The KLM has been successfully assessed in a number of novel computing environments (Manes et al., 1998, Stanton and Young, 2001, Myung, 2004, John et al., 2004a, John et al., 2004b, Luo and John, 2005). Myung (2004) evaluated the use of the technique for predicting task times for Korean-language text message entry on mobile phones. It was found that, by substituting new values for M, the KLM was a useful and reliable technique for text entry user interface evaluation. Similarly, Luo and John (2005) used KLM techniques to make task time predictions for a handheld, PDA style device. With adjustments to the model, such as the development of new operators to account for novel actions, e.g. drawing letters with a stylus, the authors were able to predict task execution time within 8% of observed results. They also discovered that, with handheld devices, the placement of system response time operators is more important than in traditional studies.

There have been notable efforts to apply KLM techniques in driving contexts. Stanton and Young (2003) assessed the KLM as one of a number of ergonomics methods to evaluate the ease with which such methods could be taught to non-ergonomics specialists. The KLM was found to be easily taught and the most reliable method. The study, involving evaluation of in-car entertainment (ICE) systems, reported good correlations between KLM predicted and observed task times. However, the results are not reported in depth. A much more detailed example of the KLM being applied in a driving context is found in Manes et al. (1998). The KLM was assessed against an empirical study of drivers performing destination entry tasks on a navigation system. The authors found strong correlations between observed and predicted times. However, they also proposed new values for a number of operators, and suggested new levels of complexity to the original method. For example, different values for second presses of a button than for initial presses. Whilst this is useful from a research perspective, such developments may potentially disturb the simplicity of the method, simplicity being one of its most attractive benefits (John, 2003).

The KLM has been discussed as a potential substitute for the running of user trials in assessing compliance with the 15-second rule ((Manes et al., 1998, Green, 1999b, Green, 2003). Such an approach goes some way to addressing the limitations of user-trial based methods, or at least those user-trial methods where there is a focus on measurements of human performance. As has been mentioned, the requirement for usable prototypes for assessment with such

methods can impinge on their usefulness in an iterative design process (Green, 2003, Salvucci et al., 2005). However, as noted in section 3.1.1, there are substantial concerns over the effectiveness of the 15-second rule. Since the use of the KLM will not address any of the previously referenced concerns, we can assume that a KLM-based 15-second rule will suffer the same limitations as the user-trial based 15-second rule.

Furthermore, a more basic task modelling approach has been employed to assess the visual demand of engaging in in-vehicle tasks (Burns et al., 2006a). This focuses on the physical aspects of keystroke modelling, e.g. button presses, scrolling, and ignore cognitive operators. As such, this technique is considered simple to apply, and demonstrates reasonable correlations with task completion times as measured using the occlusion technique (Burns et al., 2006a, ISO, 2007). However, by not considering cognitive operators, observed correlations are not necessarily as strong as they could be, and it could be suggested that the approach is overly simplistic.¹⁰

3.2.4 Using software tools to automatically generate KLMs

The failure by user interface designers to embrace GOMS and KLM techniques has been attributed, in part, to the effort that must be invested into learning how they are to be applied (John, 2003, John et al., 2004a). This may be somewhat unfair. For example, Stanton and Young (2003) demonstrated that the KLM is amongst the easiest ergonomics methods to teach and learn. Whether this limitation is founded in fact or misperception, it is disagreeable that valid and

¹⁰ Burns, P., personal communication

reliable techniques such as the KLM should be marginalised for this reason. To this end, John et al. (2004a) present a software tool that allows user interface designers to mock-up new interfaces as HTML storyboards, and demonstrate tasks on this storyboard. The software subsequently translates these actions into accurate KLMs.

Previous GOMS modelling tools have failed to gain popularity as they are not easily hooked up to system mock-ups, and often require users to learn new and complicated programming languages (Baumeister et al., 2000, John et al., 2004a). Consequently, John et al. (2004a) developed a tool that uses a what-you-see-is-what-you-get interaction metaphor for the development of mock-ups, and allows for changes in the mock-up to be automatically reflected in model predictions. The resulting tool, known as CogTool (John et al., 2004b), combines Macromedia Dreamweaver, for building interface mock-ups, the Netscape web browser, the Behaviour Recorder, for observing task procedures (Koedinger et al., 2003), and the ACT-R computational cognitive architecture commonly used for behaviour modelling. Task procedures are translated into ACT-Simple, which bears similarities to the KLM, and then translated into ACT-R in order to provide task time predictions.

A significant advantage of this tool is that it eliminates the variability in modelling predictions that are due to subjective decisions on the part of the modeller. Assuming tasks are demonstrated in the correct way, there should be no discrepancy between a novice modeller's predictions, and those of an experienced modeller, when using the software (John et al., 2004a).

Furthermore, the authors suggest that the increased cognitive modelling capabilities of ACT-R mean that resulting predictions are more accurate than comparable models made by experts. However, use of the tool may detract from the value of the traditional KLM procedure. Namely, that there is inherent merit in the modelling process itself, providing valuable qualitative information about task procedures (John, 2003, John et al., 2004a).

3.2.5 Predicting distraction through cognitive models of driver behaviour

Distract-R is a tool for prototyping and evaluating new IVIS (Salvucci et al., 2005). A cognitive model of driver behaviour in an ACT-R cognitive architecture (Salvucci et al., 2001, Salvucci, 2006), is integrated with CogTool, enabling designers to rapidly prototype interfaces, to allow those with little knowledge, or understanding, of cognitive modelling, to evaluate the distraction of various interface configurations (John et al., 2004b, Salvucci et al., 2005). There are five stages to an evaluation:

- **Rapid prototyping.** In this stage, users (IVIS designers) mock up an interface configuration, on a computer, with an emphasis on the different modalities (e.g. visual, manual, auditory) used by typical in-vehicle interfaces (Salvucci et al., 2005). Although the authors recognise that the prototyping tool is simple by the standards of comparable software (Landay and Myers, 1995, John et al., 2004a, Salvucci et al., 2005), the main objectives are achieved. It allows interfaces to be quickly sketched, with a focus on interface modality, a key consideration for developing less distracting interfaces (Salvucci et al., 2005).

- **Modelling by demonstration.** Here, users demonstrate particular tasks on the interface. These actions are translated into a cognitive model, in the ACT-Simple language, which bears similarities with KLM operators (Salvucci et al., 2005). Currently, the software does not determine the placement of cognitive operators (Salvucci et al., 2005). Furthermore, the user is required to manually specify 'switch points' in the task, where a driver may disengage from the IVIS task to monitor the driving situation (Salvucci et al., 2005).
- **Specifying individual variability.** The tool allows users/designers to specify particular driver characteristics, in order to address the issue of individual differences (Salvucci et al., 2005). Users can specify age characteristics as 'younger' or 'older', and steering aggressiveness (Salvucci et al., 2005).
- **Specifying simulation environments.** At this stage, the user determines the driving environment in which assessment will take place. Although at present only one option is available, a scenario requiring drivers to follow a lead vehicle. However, it is anticipated that future developments will range from simple car-following scenarios to multi-lane, busy roads, potentially with additional hazards (Salvucci et al., 2005).
- **Simulation and visualisation.** Here, the tool simulates instances for all the specified tasks and configurations (Salvucci et al., 2005). Ten minutes of simulated driving are run in approximately three seconds, allowing rapid evaluation of an interface (Salvucci et al., 2005). Output from the simulation includes task times, root mean squared lateral

deviations from lane centre, and, average absolute velocity (Salvucci et al., 2005).

Two validation studies have compared Distract-R predictions with predictions based on expert models and empirical results obtained during simulator trials (Salvucci et al., 2005). The first addresses the ability of Distract-R to predict the effects of interface modality on driving performance. Four configurations of a phone dialling task were conceived: *full-manual*, entering a seven digit phone number by hand; *speed-manual*, entering a phone number using a single digit shortcut; *full-voice*, using speech input to enter a whole phone number; and, *speed-voice*, entering a single digit shortcut through speech (Salvucci et al., 2005). In the empirical study, significant distraction effects were observed for the manual interfaces (Salvucci, 2001). Results for lateral deviation are shown in Figure 3.6.

As shown, both the expert model and Distract-R rank the different conditions in a similar fashion to the empirical results. The authors are encouraged by the results, as Distract-R predicts the large effect for the *full-manual* condition (Salvucci et al., 2005). In the figure, however, we see that, whilst Distract-R predictions are generally smaller than observed results, for the *full-manual* condition the prediction is substantially greater. The authors do not address this issue in their discussion (Salvucci et al., 2005).

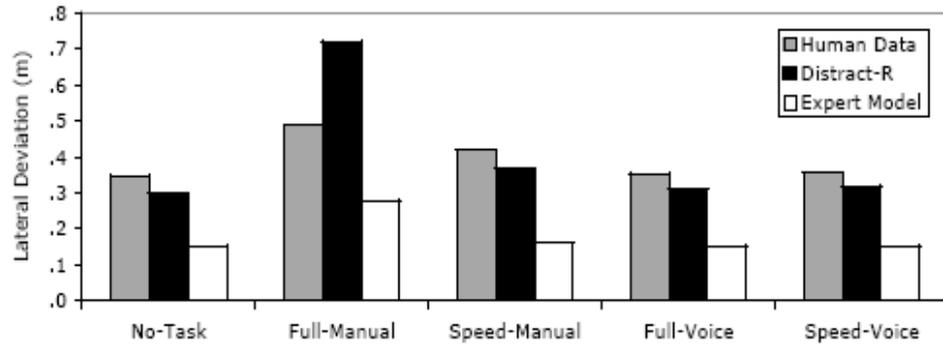


Figure 3.6 Lateral deviation results for input modality validation study (Salvucci et al., 2005)

In a second study examining age effects, Distract-R predictions are compared with results from an empirical study carried out by Reed and Green (1999). Younger and older drivers were compared when driving in two conditions: (1) no task, (2) whilst performing a 10-digit phone-dialling task (Reed and Green, 1999, Salvucci et al., 2005). Empirical results demonstrate that there is little difference between performances when there is no secondary task (Reed and Green, 1999, Salvucci et al., 2005). However, the presence of a task had an effect on driving performance, and this effect was larger for older drivers (Reed and Green, 1999, Salvucci et al., 2005). Comparison results with Distract-R and expert model predictions (Salvucci et al., 2004) are shown in Figure 3.7.

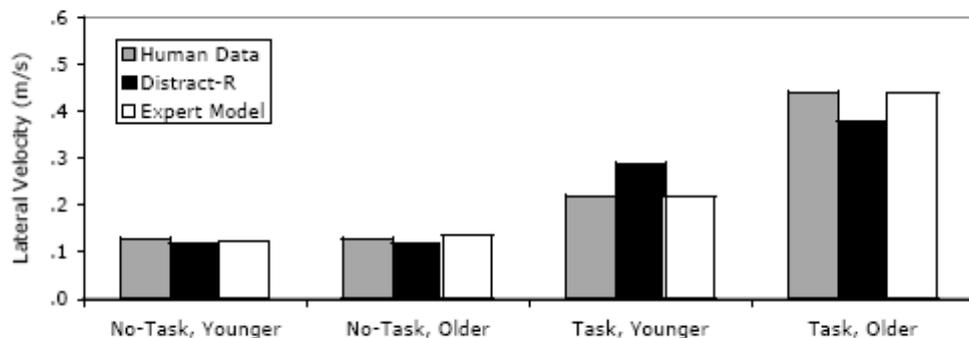


Figure 3.7 Lateral velocity results for age effects study (from Salvucci et al., 2005)

Both prediction methods accurately show no age effect in the no-task condition (Salvucci et al., 2005). Furthermore, both demonstrate an effect on driving

performance of performing the task, and the corresponding age interaction (Salvucci et al., 2005). However, the effect and interaction is more closely anticipated by the expert predictions (Salvucci et al., 2005). Despite this, the authors conclude that Distract-R accounts for the basic effects well, in both a quantitative and qualitative fashion (Salvucci et al., 2005).

Distract-R is potentially a very interesting tool for driver distraction research. However, there are some causes for concern. In particular, the tool requires designers to determine appropriate switch points where it is likely that drivers will break from an in-vehicle task to monitor the driving environment. Currently, the theory behind task switching is poorly understood, and an area of ongoing research (John et al., 2004b). Furthermore, there does not appear to have been any independent validation of the technique.

3.3 Opportunities for research

Whilst new and emerging IVIS technologies have the potential to increase the safety, comfort and efficiency of the driving experience, this must be balanced with design that focuses on minimising the potential for drivers to be distracted. Thus far, driver-centred design processes have employed, and adapted, many of the same global evaluation techniques common to user-centred design practice. The most effective, or at least the best established, of these techniques are assessed during performance-focussed user trials. The effectiveness of a range of methods, in terms of validity and economy, is presented in Figure 3.8. It must be stressed that the placement of the various methods within this figure is not as a result of a quantification of their

respective economy or validity, it is merely a representation of the impression given by the literature support for each. Similarly, the representation of opportunities for research is a personal interpretation and its boundaries should be viewed as being reasonably loose. It could be argued, for example, that this shaded region would best appear in the top left of the figure – low cost and high validity – although such a placement would be perhaps too idealistic.

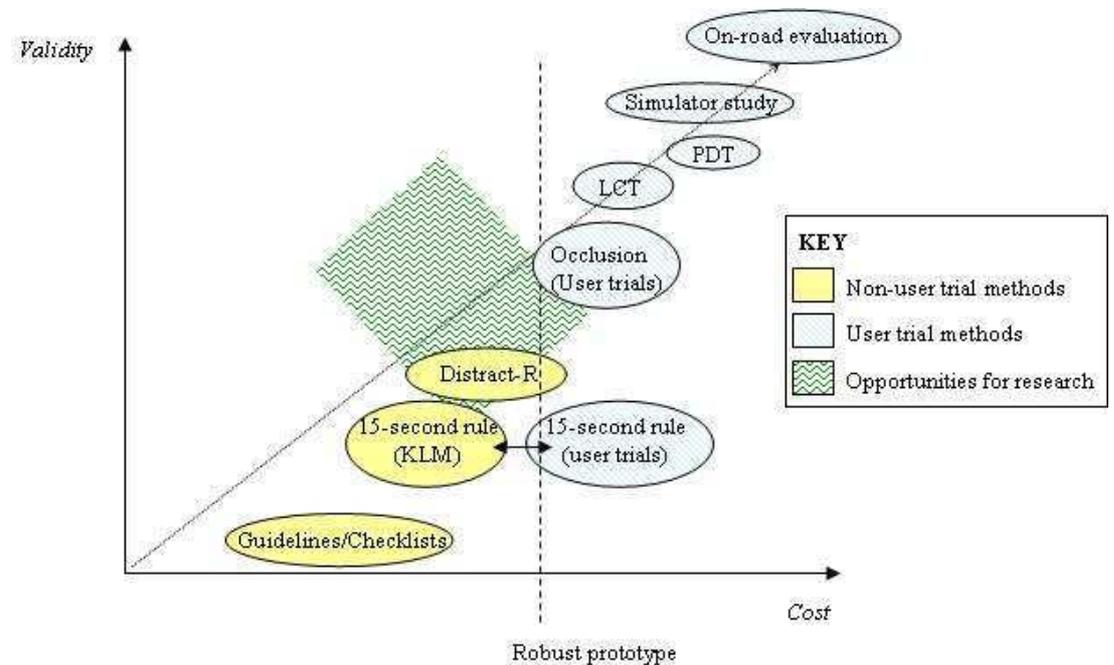


Figure 3.8 Evaluation methods by economy and validity

Methods may be further categorised according to their reliability. In this respect, the methods reviewed in this chapter might be placed within the framework given in Chapter 1 (Figure 1.2) as shown in Figure 3.9. Similar to the previous figure, the placement of methods here is somewhat subjective. However, the figure demonstrates that few methods appear close to satisfying the three criteria of validity, economy and reliability. An exception is Distract-R, which potentially satisfies all three conditions. However, it is placed only

on the edge of the validity construct in this case since there has so far been a lack of investigation by the wider research community, particularly within transport and safety research departments to the author's knowledge. Further, a concern over the way the technique handles task-switching has been raised, as noted earlier (John et al., 2004b).

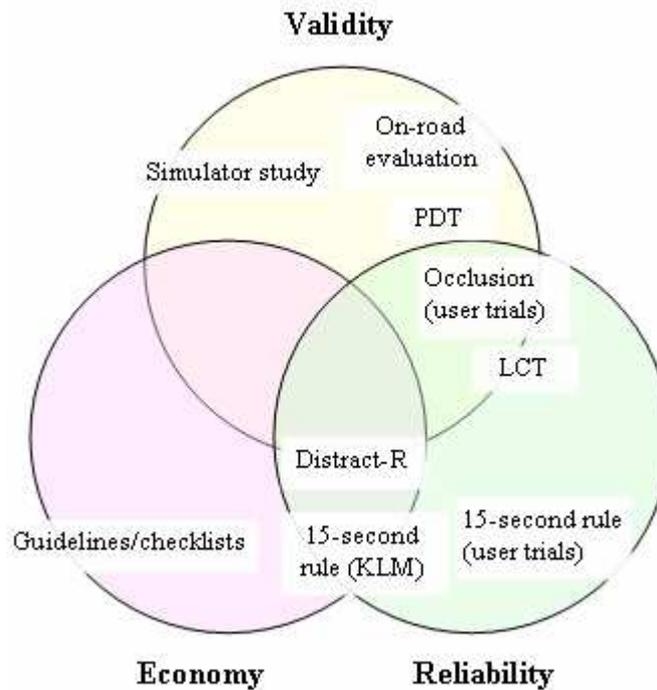


Figure 3.9 Evaluation methods according to validity, economy and reliability

Performance focussed user-trials share a common limitation; in order to be effective they require a robust, functional prototype (Green, 2003, Salvucci, 2005, Angell et al., 2006). Efforts have been made to improve the economy of running trials, e.g. an occlusion study is significantly more cost effective than an on-road evaluation. However, within the wider, practical, context, of an IVIS development cycle, all user trial methods fall towards the higher end of the economy scale. In order to overcome this limitation, there is a need for

research that focuses on the development of non-user trial based methods, which do not compromise on the validity of results.

In order to bridge the gap between non-user trial techniques and truly economical methods, it will be necessary to find appropriate substitutes for the rich information provided by user-trials. Models of human performance may provide such a solution. Indeed, the 15-second rule, which was originally developed as a user-trial technique, has been assessed using KLM task time predictions (see section 3.2.2). Ultimately, the validity of making distraction assessments on task time only has been called into question, and the effects of these misgivings are felt regardless of the assessment protocol followed.

If human performance in another user-trial method can be modelled, the resulting technique would retain the validity of its user-assessed counterpart, whilst improving on its economy. In this respect, the occlusion technique appears to be the most suitable candidate. Similar to the 15-second rule protocol, the technique is a single-task method (Angell et al., 2006), with its focus on the performance of IVIS tasks; other techniques tend to focus on the effect of IVIS task performance on other variables. In demonstration of the similarities between the techniques, it has been proposed that the 15-second rule protocol be extended to include an occlusion component, in order to satisfy misgivings regarding the rule's effectiveness (Green, 2003). Therefore, it can be assumed that the process of modelling human performance in occlusion trials will be more straightforward than modelling performance in the LCT, for example. Burns et al. (2006a) attempted to predict visual demand

through a simplified keystroke-level modelling procedure, counting inputs. Correlations were observed between task times in occlusion conditions and model predictions, although these weren't as strong as hoped. However, the results suggest that keystroke modelling and the occlusion technique are not incompatible.

Development of a modelling technique for the occlusion protocol is dependent on a mixture of empirical and non-empirical work. Whilst the validity of the occlusion technique has been demonstrated in previous studies, there is still opportunity for further assessment; in particular, focussing on the validity of the method as set out in the ISO standard and in comparison with results from on-road studies. Similarly, there is scope for further assessment of the application of the KLM within the driving context. In addition, such assessments will provide practical knowledge of both techniques, in order to support the development of a hybrid technique as discussed above. There will subsequently be a requirement to validate a hybrid technique, against empirically generated results from an occlusion study. A final stage in the development of the new technique will be an assessment of its reliability.

3.4 Summary

This chapter has presented a review of the available evaluation methods, user-trial and non-user-trial based, for evaluating IVIS in terms of their potential distraction. Methods are both qualitative and quantitative in nature, and can be implemented across the IVIS design cycle. However, the most mature methods generally fall into the latter stages of design, and this is considered a limitation.

It is suggested that such methods are not truly effective in a driver-centred design approach; since it is unlikely their results can be translated into meaningful design changes, due to practical, economic considerations (Green, 2003, Salvucci et al., 2005). Therefore, it is suggested that research should focus on the development of a modelling technique that seeks to predict measures normally obtained through occlusion trials.

4 Study 1 – Validation of the occlusion technique

The study outlined in this chapter assesses the validity of the occlusion technique, introduced in the previous chapter as a potentially useful method for evaluating the potential visual demand of an in-vehicle device. An on-road evaluation study is discussed, in which participants completed two tasks on two different systems. Results of this study are compared with results from an occlusion assessment of both tasks and systems.

Chapter 3 presented a review of literature regarding the occlusion technique. It emerged that the majority of studies have aimed to validate the occlusion technique using data from simulator trials. Few, in contrast, have compared occlusion results with on-road studies (a notable exception is Baumann et al., 2004). By measuring driving performance variables in a real vehicle, driving on public roads, participants' behaviour will inevitably be more naturalistic than when the same tasks are performed in a simulator (Reed and Green, 1999, Young et al., 2003).

Furthermore, this study aims to evaluate the occlusion technique as a design tool enabling a researcher/developer to make design decisions regarding alternative user interfaces. Do results from occlusion trials reflect participants' on-road performance, and, if so, is this information a better reflection of the visual demand of an IVIS than simply measuring static task time? Consequently, the study will consider the use of the occlusion technique to compare between interfaces, and, as a technique for assessing conformity with

cut-off values, e.g. the 15 second TSOT limit suggested by AAM (AAM, 2003).

In addition, the study provides an opportunity to examine the R metric in comparison with on-road measures of visual behaviour. The metric has been described as a measure of how easily a task is resumed following an interruption; an assumption might be that IVIS tasks with higher R-values may encourage longer single glances to a system whilst driving, for example. Finally, the study provides an opportunity to assess the draft international standard and report on any procedural difficulties encountered.

4.1 Research Questions

This study seeks to answer the following questions:

1. Does the occlusion technique accurately assess the visual demand of IVIS in comparison with measures of visual behaviour taken from an on-road evaluation?
2. Do results from an occlusion study provide any more information regarding potential visual demand than assessments based on static task time?
3. How effective is the technique as a comparison tool for assessing different IVIS task designs, and, in assessing compliance with cut-off values?
4. What information is provided by the R metric?
5. What are the benefits or problems associated with the occlusion technique?

4.2 *Research contribution*

In addressing the above research questions, this study contributes to knowledge in the following ways:

1. It is a rare example of comparing occlusion results with measures taken from an on-road study, with participants driving in real traffic
2. It examines the occlusion standard in a practical manner, and in keeping with the spirit of the method: e.g. to evaluate developing in-vehicle systems. Previous studies have not followed the standard protocol, and have utilised existing systems.
3. It allows for an examination of the R metric in relation to on-road measures of driver distraction. At the time it was conducted, few authors have addressed the specific characteristics of the measure in the published literature.

4.3 *Method*

4.3.1 *Participants*

Sixteen licensed drivers took part in the trials. The mean age was 41.2, SD=19.8, range 22 to 68, with an even gender split. Driving experience was noted by annual mileage (see Table 4.1 for a summary). Participants had held full driving licenses for an average of 22 years (SD=19, range 5 to 54) and were paid for their time.

Table 4.1 Participant driving experience

Annual Mileage	Participants
0 to 1 000	1
1 001 to 5000	3
5001 to 10 000	7
10 001 to 15 000	2
15 001 to 20 000	2
20 000 +	1

4.3.2 Equipment

Trials were conducted in a Honda Civic. Static and occlusion trials were performed in the stationary vehicle. Two systems were assessed in the study. System A incorporated a commercially available, integrated navigation system/entertainment system with the vehicle's original heating, ventilation and air-conditioning (HVAC) controls. System B was a prototype system, which combined navigation, ICE and HVAC controls, which had been designed with the intention of minimising visual demand by affording non-visual ('blind') interaction. The trial car was equipped with four cameras to record: the driver's face; the road scene ahead; the system controls and display; and, the overall driver cockpit environment (see Figure 4.1). In the occlusion trials, PLATO (portable liquid-crystal apparatus for tachistoscopic occlusion) goggles were used as a means of achieving occlusion (Milgram, 1987, Milgram, 2005). These goggles are commonly used for that purpose (Stevens et al., 2004). A laptop computer was used to alternate the translucency/transparency of the goggle lenses at the regular intervals required by the version of the draft standard as available at the time (ISO, 2004b).



Figure 4.1 Camera views

4.3.3 Tasks

Two tasks were selected for the trials based on the comparability between systems, their suitability for occlusion, and the limitations of the prototype system. In particular, the ISO standard requires that tasks assessed by the occlusion technique have static task times of greater than five seconds (ISO, 2007). The first task involved the navigation function of each system, and required participants to select a destination from a stored list and confirm it (navigation task). The second task involved manipulation of the sound characteristics within the vehicle, such as adjusting the levels for treble, bass, balance and fade (sound task). Both tasks could be satisfactorily completed on each system, and on average performance times exceeded five seconds.

4.3.4 Trial set-up and procedure

The study design was the result of consultation between members of the BIONIC project team. Several pilot studies were conducted to refine the

procedure, and the method was ethically approved by Loughborough University.

Prior to the occlusion and road trials, participants were made familiar with the car and its controls and fully trained in the use of the system. In a repeated measures design, each participant completed two trials, one for each system, with a delay of around two to four weeks between trials. This was felt to minimise any learning effects regarding trial procedure. To avoid order effects, half of the participants completed the trials with the order System A followed by System B, whilst the other half followed the order System B followed by System A.

The occlusion trials were conducted whilst sat in the stationary vehicle, prior to road trials. The procedure for the occlusion trials followed that given in the most up-to-date version of the draft standard available at the time of the trials (ISO, 2004b). Participants were trained in the tasks and given five practice trials in order to familiarise themselves with the task/system. An assessment was made of their ability to perform the tasks to a satisfactory level before the full occlusion trial. Five task instances were undertaken with full vision and five under occlusion conditions, as per the draft standard. Ordering of static versus occluded trials was counterbalanced to minimise learning effects.

Road trials took place on a section of the A50 in Leicestershire/Derbyshire, United Kingdom; a two-lane dual carriageway restricted by the national speed limit of 70 mph. Participants were instructed to try to maintain a speed of 70

mph throughout the trial, traffic permitting. During this section of the trial, participants drove a 20-minute route twice whilst performing several tasks, including the two examined in the earlier occlusion trials. Video data was captured on the second trip, by which time it was felt participants were fully competent in performing the tasks whilst driving.

4.4 Results

A number of technical issues were encountered during the trials, meaning that data was not gathered from the full sample of 16 participants. The quantity of participant data collected under each condition is detailed in Table 4.2.

Table 4.2. Participant data numbers for each condition

	Static task time		Occlusion measures		Road trial measures	
	<i>Navigation</i>	<i>Sound</i>	<i>Navigation</i>	<i>Sound</i>	<i>Navigation</i>	<i>Sound</i>
	<i>Task</i>	<i>Task</i>	<i>Task</i>	<i>Task</i>	<i>Task</i>	<i>Task</i>
System A	14	15	15	15	13	14
System B	14	15	12	11	14	14

4.4.1 Results from occlusion trials and on-road evaluation

The results of the occlusion trials and the on-road evaluation are presented in Tables 4.3 and 4.4. Under the terms of the 15-second rule, assessing the systems and tasks according to the static task time results, it can be seen that both systems are acceptable for the navigation task, although only System A passes for the sound task. It should be noted that the procedure involved in measuring static task time in this study is not the same as that recommended by the 15-second rule. However, similar to the occlusion technique, the 15-second rule involves participants performing several variations of a task and obtaining mean task time values (SAE, 2000). Indeed, the procedure for collecting static task times according to the occlusion standard could be described as more

rigorous than the 15-second rule, therefore any differences between the protocols are not considered problematic.

According to the AAM limit of 15 seconds for TSOT, all tasks and systems would be considered acceptable (AAM, 2003). The resumability ratio, R , in all cases, is beneath the indicative guideline value of 1 as suggested by the ISO standard (ISO, 2007), with significantly lower values for the sound task compared to the navigation task (System A: $t(21)=2.01$; $p<0.05$. System B: $t(18)=4.83$; $p<0.0005$). A significant difference was also detected between the systems for the sound task, with System B significantly lower than System A ($t(24)=3.08$; $p<0.01$). There was no significant difference detected between the systems for the navigation task. In the on-road results, the key measures are single glance duration and total glance time. For single glance duration, no significant differences were observed between any of the tasks or systems under investigation. All systems and tasks had mean single glance durations of less than 1.5 seconds. For total glance time, only one significant difference was detected, with System A showing a significantly lower value for the sound task than System B ($t(24)=-2.3$; $p<0.05$).

Interpreting the results from the occlusion study, we would expect that all tasks should be completed satisfactorily whilst driving, according to the suggested limits of 15 seconds for TSOT (AAM, 2003) and a value of R less than 1 (ISO, 2007). It would be reasonable to expect, however, that the sound task on both systems would be less visually demanding than the navigation task, given the lower R -values for these tasks. There are indications that this may be the case

within the results from the on-road study. For System A, the number of glances made to the system during the sound task compared with during the navigation task is approaching significance ($t(23)=2.01$; $p=0.056$). For System B, a smaller percentage of the dynamic task time was spent glancing to the system, which also approaches significance ($t(26)=2.05$; $p=0.05$). Given a significant difference was detected between the systems for the R-value for the sound task, some indication that there is a difference in the visual demand between the systems for this task in the on-road results might be expected. In this case, a significant difference was detected in the proportion of dynamic task time spent glancing to the system ($t(26)=2.17$; $p<0.05$).

Table 4.3 Comparison between systems for results from occlusion and on-road trials, showing mean values with standard deviations in brackets

		Occlusion			On-road		Proportion	Glance	Single Glance
		STT (s)	TSOT (s)	R	DTT (s)	TGT (s)	TGT/DTT (%)	Frequency	Duration (s)
Navigation Task	System A	11.0 (4.36)	9.3 (2.75)	0.94 (0.33)	21.4 (10.37)	12.7 (5.90)	61.8%	11.5 (3.91)	1.08 (0.20)
	System B	11.4 (3.11)	9.4 (1.78)	0.84 (0.20)	22.0 (8.97)	12.9 (5.26)	60.7%	11.1 (3.93)	1.17 (0.24)
Sound Task	System A	13.2** (3.90)	9.1** (3.73)	0.67 (0.17)	16.4** (6.60)	9.8* (4.53)	60.8%	8.8** (3.12)	1.10 (0.21)
	System B	26.4 (5.98)	13.9 (3.32)	0.51** (0.12)	30.4 (11.21)	14.4 (5.91)	49%*	12.7 (4.05)	1.12 (0.31)

STT = static task time, TSOT = Total Shutter Open Time, R = resumability ratio (TSOT/STT), DTT = dynamic task time, TGT = total glance time.

* Indicates value significantly lower than for opposing system ($p < 0.05$). ** Indicates value significantly lower than for opposing system ($p < 0.01$).

Table 4.4 Comparison between tasks within systems for results from occlusion and on-road trials, showing mean values with standard deviations in brackets

		Occlusion			On-road		Proportion	Glance	Single Glance
		STT (s)	TSOT (s)	R	DTT (s)	TGT (s)	TGT/DTT (%)	Frequency	Duration (s)
System A	Navigation task	11.0 (4.36)	9.3 (2.75)	0.94 (0.33)	21.4 (10.37)	12.7 (5.90)	61.8%	11.5 (3.91)	1.08 (0.20)
	Sound task	13.2 (3.90)	9.1 (3.73)	0.67* (0.17)	16.4 (6.60)	9.8* (4.53)	60.8%	8.8 (3.12)	1.10 (0.21)
System B	Navigation task	11.4** (3.11)	9.4** (1.78)	0.84 (0.20)	22.0* (8.97)	12.9 (5.26)	60.7%	11.1 (3.93)	1.17 (0.24)
	Sound task	26.4 (5.98)	13.9 (3.32)	0.51** (0.12)	30.4 (11.21)	14.4 (5.91)	49%	12.7 (4.05)	1.12 (0.31)

STT = static task time, TSOT = Total Shutter Open Time, R = resumability ratio (TSOT/STT), DTT = dynamic task time, TGT = total glance time.

* Indicates value significantly lower than for opposing task ($p < 0.05$). ** Indicates value significantly lower than for opposing task ($p < 0.01$).

4.5 Discussion

The occlusion technique has been developed as a method to assess the potential visual demand of IVIS tasks. Methods such as the 15-second rule, which makes assessments based on static task time, are felt to be problematic, being unable to discriminate between tasks that require long glances and those that can be effectively completed with a series of short glances (Young et al., 2003, Stevens et al., 2004). The technique was used to assess two tasks on two IVIS, which were subsequently investigated as part of an on-road evaluation study. The purpose of the study was to assess the validity of the technique and to evaluate its ease of implementation.

4.5.1 Validity of technique

Measures obtained from the occlusion trials reflected those captured from the on-road evaluation. Results for TSOT were mirrored in on-road measures of dynamic task time and total glance time. The AAM's suggested limit of 15 seconds for TSOT means that both tasks on both systems were acceptable according to this measure. This is in contrast to the static task time results, where, in accordance with the 15-second rule, the sound task on System B would have been deemed unacceptable. Results from the on-road evaluation do not necessarily suggest that this task is inherently more visually demanding than the other tasks under investigation, with single glance durations unaffected in comparison with the other task/system. When compared with the navigation task on System B, no significant difference was detected in total glance time. There have been suggestions that a criteria of TSOT < 15 seconds is a more useful measure than the 15-second rule, which has the potential to

'fail' tasks that are not necessarily unsafe to perform whilst driving (Tijerina et al., 2000). The results reported here appear to support this claim.

The second measure to be obtained from the occlusion trials is the resumability ratio, R . A value of R greater than one is said to indicate that participants may be having difficulty completing the task in the brief periods of vision available to them during occlusion trials (ISO, 2007). The expectation may be that this should be reflected in the measure of single glance duration, with greater R -values corresponding to longer single glance durations. Such a relationship was not shown in the results from this study. However, there is some indication that lower R -values may indicate reduced visual demand. For System A, significantly fewer glances were made to the system in the sound task, which had an R -value of 0.67, in comparison with the navigation task, where the R -value was 0.94. Glance frequency has been proposed as a more important indicator of visual demand than single glance duration (Rockwell, 1988). Further, for System B, the proportion of total glance time to dynamic task time was significantly smaller for the sound task ($R=0.51$), than for the navigation task ($R=0.84$).

The expectation that R should relate to single glance durations is flawed, as it does not recognise that R is influenced by two main factors. Namely, how much of a task can be completed in brief periods of vision and how much can be done whilst deprived of vision, also termed blind operation (Burns, 2006). Hence, a low value of R may be the result of a task where participants were able to significantly further their progress in the task during occluded periods.

The results of this study, and their interpretation, suggest that R is not quite the elegant measure it was hoped to be. However, that does not mean it is without merit. Care should be taken to fully understand the factors that influence R . In particular, it is recommended that tasks where R exceeds one be treated with caution. In such cases, it is suggested that the costs of resuming a task following interruptions are greater than, or negate, any potential for blind operation (Burns, 2006).

An analysis of visual behaviour for tasks with an R -value greater than 1 may clarify the position as to whether R can truly be described as a measure of visual demand. Such tasks would be primarily visual in nature, with little opportunity for blind operation. Noy et al. (2004) reported $R > 1$ for two simulated visual search tasks, where participants were required to locate a three letter stock code amongst a list of other codes, and report the associated price value. Similarly, Angell et al. (2006) report just one task with mean R -values greater than one, a task that involved reading complex map information from a display. An R -value greater than one would suggest that task performance is adversely affected when visual attention is restricted to short glances, requiring drivers to invest extra visual effort reorienting to tasks following interruptions, and it is anticipated this may lead to longer single glance durations in a driving environment.

4.5.2 Ease of implementation

There are few quantitative techniques that IVIS developers may use in early design stages. The occlusion technique offers significant benefits compared with traditional methods that rely on simulator or road studies, particularly in

terms of cost and procedure. Occlusion trials do not require expensive simulators or instrumented vehicles, and the analysis of data is much more efficient. However, the occlusion technique is a method based on measures of human performance gathered in user trials, and this has associated problems for its application during earlier design stages (Green, 2003, Salvucci et al., 2005).

To accurately measure human performance prototype systems must be robust and reliable, in order to withstand the rigours of user trials. However, such prototypes are generally only feasible in the later stages of design. Correspondingly, it may be uneconomical to make significant changes to a design at this point (Green, 2003). As a case in point, one of the systems used in the study reported here was a prototype, and was found to be unreliable, causing frustrations for both participants and experimenters; whilst many changes were proposed to the system during pilot studies, these were seldom applied due to time and financial constraints. As such, the occlusion technique may be more useful as a tool for deciding what tasks are acceptable for on-the-move access, and which should be ‘locked out’ whilst the vehicle is in motion. Correspondingly, its effectiveness as a design tool is likely to be limited.

The technique is relatively easy to implement, although it is more complicated than simply measuring static task time. This is mostly due to the need to familiarise participants with the occlusion procedure. In contrast, on-road studies can be highly problematic: they are costly to run and time consuming; real-world traffic and environmental conditions are unpredictable; and, there are greater ethical considerations when asking participants to engage in

activities that are potentially unsafe (Young et al., 2003). Furthermore, the analysis of data from road trials can be very time consuming.

4.6 Summary

In summary, the occlusion technique was found to be a valid method for evaluating the visual demand of IVIS tasks. The findings presented in this chapter support claims that the TSOT measure is more sensitive to visual demand than the 15-second rule, being able to distinguish between tasks based on the amount of visual control required to achieve a goal. The study highlights issues regarding the R metric, namely that it is influenced both by how far a task may progress in vision periods, and how much can be done in non-vision periods, although the metric is not considered completely redundant. The experience of running the trials confirms that many of the procedural issues faced by other performance-based user trial methods remain. Whilst the dataset may be easier to analyse, quickly producing results, in comparison to other methods, the need to run trials on reliable, robust prototype systems casts doubt on its effectiveness within a formative design cycle.

5 Study 2 – Applying the keystroke level model in a driving context.

The previous chapter assessed the validity of the occlusion technique as a method for evaluating the visual demand of an IVIS. Whilst it was concluded that there is merit in the technique, some limitations were identified. In particular, the requirement to evaluate systems through user trials, with a reliance on measures of performance, means that well-developed prototypes are necessary for successful occlusion trials. As such, in practical terms, the occlusion technique (as with all performance based, user-trial techniques) is unlikely to be used within an iterative design cycle as intended; namely, to inform redesign.

To overcome this issue it is necessary to investigate alternative, quantitative methods that are less reliant on performance based user trials. The family of GOMS techniques, introduced in Chapter 3, particularly the keystroke level model (KLM), are of interest here. The KLM was originally developed for desktop-computing scenarios (Card et al., 1983), and has since been proven reliable in this, and other, contexts (see Chapter 3 for a fuller discussion). Assessments of the technique have previously been made in in-car environments with IVIS. Manes et al. (1998) outlined the development of a number of operators, and operator values, specific to in-car navigation systems (see also Green, 1999b, Green, 2003). Work by Stanton and Young (2003) suggested that the technique is reliable and valid in a driving context, however there is a lack of specific detail within the results presented. For example, their

study considered a range of ergonomic methods and the effectiveness of these when applied by non-ergonomics professionals. As such, the study does not report key information of relevance here, such as observed and predicted task times, error rates, and so on, that would be expected from a thorough analysis of the KLM. Therefore, a study considering application of the technique using standard operator values for in car systems, and reporting detailed results is deemed a useful contribution to research.

KLM predictions of static IVIS task time have been proposed, and assessed, as a surrogate for the running of user trials when evaluating IVIS based on the 15-second rule (Manes et al., 1998, Green, 1999b, Green, 2003). Due to the various perceived shortcomings of this rule, this study does not intend to validate the conventional KLM as a technique for assessing the potential distraction posed by an IVIS. Moreover, the study is a further exploration of the technique's portability to novel contexts and its ability to predict task times for devices other than desktop computing.

Finally, the experience of running this study will provide sound practical knowledge regarding application of KLM techniques. This experience will be essential in the development of more complicated versions of the KLM that may allow modelling of human performance under alternative conditions, for example occlusion.

5.1 Research questions

The question addressed by this study is as follows:

1. Is the KLM a valid substitute for empirical assessments of static IVIS task time performance?

5.2 Research contribution

This study contributes to research by examining the KLM in its traditional form as a method for predicting performance times for tasks on in-car radio systems. Previous studies have either not reported their results in depth (e.g. Stanton and Young, 2004) or have utilised a range of different operator values and with less generic systems (e.g. Manes et al., 1998). Further, KLM studies generally tend to focus on making comparisons between tasks on individual systems, whereas this study considers its use in comparing tasks on multiple systems.

5.3 Method

An empirical study was conducted to assess the applicability of the KLM in a driving context. User trials were conducted with twelve participants (mean age= 24; range: 22 to 28; 50:50 gender split) on two in-car entertainment (ICE) systems in a counterbalanced order. System 1 was a Fiat Punto 2001 Blaupunkt Radio; System 2 was a Citroen ZX Avantage 1991 Standard Citroen Radio. System 2 was ten years older than System 1, and the interaction with the device was more mechanical in nature.

For each system, eleven tasks were performed in a fixed order (see Table 5.1). Each user was fully trained in all task procedures before the trials. Training consisted of verbal instructions, followed by the opportunity to practice various tasks, and a ten-minute familiarisation period. As KLM predictions are associated with expert performance, it was required that participants were suitably skilled in the use of the ICE systems. Consequently, an assessment of each participant's competence was made prior to recording actual data. Trials took place in stationary vehicles, and participants were filmed performing the tasks to ensure that accurate task times were recorded. Trial procedure was piloted before actual trials took place.

Table 5.1 Tasks

Task no.	Task	Task no.	Task
1	Switch on tuner	7	Decrease volume
2	Tune into Radio1 (FM 97.9)	8	Insert cassette
3	Save frequency to pre-set 1	9	Increase base to full
4	Increase volume	10	Eject cassette
5	Change station to BBC Nottingham (FM 103.8)	11	Switch off radio
6	Save frequency to pre-set 2		

KLM models of each task were created by an MSc student, Darius Karbassioun (Karbassioun, 2004) according to the method set out by Card et al. (1983), which were checked, verified and corrected where necessary by the author of this thesis, and task times were predicted. Tasks were decomposed into their basic options: i.e. movements between the steering wheel and device (Reach far – RF), pressing buttons (K). Thus, for System 1, Task 1 becomes: H + K + H. Subsequently, cognitive operators (M) were added to the models, following the heuristics described by Card et al. (1983). Time values for M (1.35 seconds) and K (0.2 seconds) were drawn from Card et al. (1983), whilst Green's (2003) 'Reach far' value of 0.45 seconds was used for movements

between the steering wheel and the devices (Rf). Therefore, the model now reads: $M (1.35) + H (0.45) + K (0.2) + H (0.45) = 2.45$ seconds.

5.4 Results

5.4.1 Correlations between observed and predicted task times

Correlations between observed task times and KLM predictions were high (see Figures 5.1 and 5.2). The user trials did, however, show some small differences in task times that were not predicted by KLM. For System 2, in particular, there was greater variability with user trials in the shorter tasks, although the lengthier tasks were well correlated with the KLM predictions. For all tasks, the system with the shorter task time was accurately predicted by the KLM (see Table 5.2). The difference in value between the systems was less well predicted, however, with the KLM generally more conservative. This may be due to the relatively small sample group in the user trials, and is possibly further influenced by the use of operator values from the original KLM, which were based on desktop computing actions (Card et al, 1983).

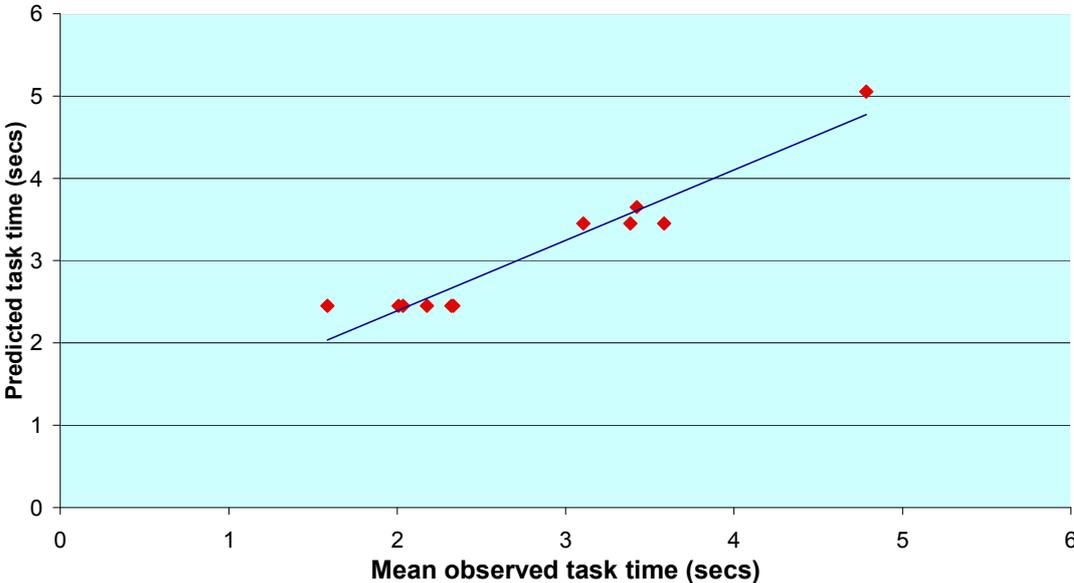


Figure 5.1 Comparison of mean observed task time and predicted task time for System 1 ($r=0.97$)

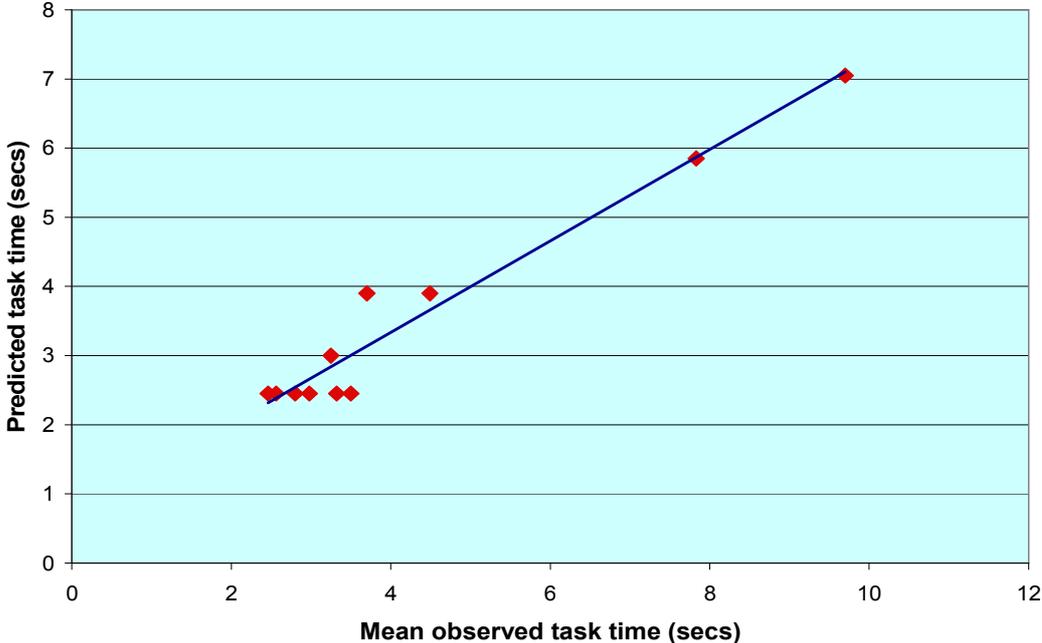


Figure 5.2 Comparison of mean observed task time and predicted task time for System 2 ($r=0.98$)

Table 5.2 Results

Task	System 1		System 2			Preferred system	
	Mean time (s)	observed (SD)	Predicted time (s)	Mean time (s)	observed (SD)		Predicted time (s)
Switch on radio	2.18	(0.61)	2.45	3.25	(1.06)	3	1
Tune to Radio 1	3.42	(0.30)	3.65	7.83	(1.49)	5.85	1
Save frequency to preset 1	3.11	(0.72)	3.45	4.49	(0.68)	3.9	1
Increase volume	2.32	(0.25)	2.45	2.80	(0.46)	2.45	1
Tune to Radio Nottingham	4.79	(0.55)	5.05	9.70	(0.85)	7.05	1
Save frequency to preset 2	3.39	(0.50)	3.45	3.70	(0.25)	3.9	1
Decrease volume	2.33	(0.46)	2.45	3.32	(0.89)	2.45	1
Insert cassette	1.59	(0.40)	2.45	2.98	(0.80)	2.45	1
Increase bass to full	3.58	(0.41)	3.45	3.50	(0.66)	2.45	2
Eject tape	2.04	(0.30)	2.45	2.46	(0.30)	2.45	1
Switch off radio	2.01	(0.29)	2.45	2.56	(0.33)	2.45	1

5.4.2 Error rates

Table 5.3 Prediction error rates for all tasks on both systems (calculated as follows: difference between 100 and predicted task time as a percentage of actual task time, consistent with John et al., (2004a))

Task	System 1			System 2		
	Actual task time (s)	Predicted task time (s)	Error (%)	Actual task time (s)	Predicted task time (s)	Error (%)
Switch on tuner	2.18	2.45	12.59	3.25	3	- 7.69
Tune into Radio1 (FM 97.9)	3.42	3.65	6.66	7.83	5.85	- 25.33
Save frequency to pre-set 1	3.11	3.45	11.08	4.49	3.9	- 13.16
Increase volume	2.32	2.45	5.56	2.80	2.45	- 12.59
Change station to BBC Nottingham (FM 103.8)	4.79	5.05	5.54	9.70	7.05	- 27.33
Save frequency to pre-set 2	3.39	3.45	1.92	3.70	3.9	5.55
Decrease volume	2.33	2.45	5.02	3.32	2.45	- 26.20
Insert cassette	1.59	2.45	54.38	2.98	2.45	- 17.65
Increase base to full	3.58	3.45	- 3.74	3.50	2.45	- 29.96
Eject cassette	2.04	2.45	20.39	2.46	2.45	- 0.28
Switch on radio	2.01	2.45	22.01	2.56	2.45	- 4.15
Average	2.795	3.07	+/-13.5	4.234	3.49	+/-15.4

KLM studies commonly report error rates as a measure of the technique's validity, in addition to correlations. It has been suggested that, when KLM

rules are applied rigorously and consistently, a modeller can expect to produce prediction times well within 20% of that taken by skilled participants (John et al., 2004b). In this study, the error rates were within this boundary, at an average for System 1 of 13.5% and 15.4% for System 2. Absolute error rates are listed in Table 5.3.

5.5 Discussion

The strong correlations between the predicted task times and observed task times are encouraging. They corroborate other research, which has suggested that KLM techniques can be applied to computing environments that do not fit the desktop paradigm for which they were designed (Manes et al., 1998, John et al., 2004a, John et al., 2004b, Myung, 2004, Luo and John, 2005). Strong correlations suggest the KLM provides good relative predictions of task times for in-vehicle tasks. However, to determine the absolute validity of the predictions, it is necessary to consider error rates. For System 1, the average prediction error rate was 13.5%, within the reasonable expectation of 20% for KLM predictions (John et al., 2004b). When the insert cassette task, which had the greatest error rate, is removed the figure changes to 9.45%. It is probable that the keystroke operator (K) used to represent the insertion of the cassette does not accurately reflect real life task performance due to differences in the physical action, thus accounting for the greater variation between observed and predicted times.

This finding may explain why the KLM generally under predicted for System 2. Since the correlation between predicted and observed results was equally

strong for both systems, it is possible that the weaker error rate for System 2 can be attributed to inconsistencies between the operator values used and actual performance. Put simply, the time to execute actions on System 2 was not consistent with the operator time values used in the models. This is a plausible explanation; System 2 was an older generation ICE system, with a more mechanical interface. Operator values for pressing buttons, for example, were originally formulated following observations of users of keyboards with desktop computers. Button presses on System 2 are therefore likely to have taken participants longer than a K operator assumption of 0.2 seconds.

The nature of the tasks and the ICE interfaces meant that compromises had to be made when modelling performance. The KLM operators used were drawn from the original pool of operators outlined in Card et al. (1983) with a driving specific operator for movement from steering wheel to device taken from Green (2003). These operators were developed for desktop computing tasks, and comprise of actions such as keystrokes, movements between the keyboard and mouse, and so on. There was no available operator for actions such as turning a knob, e.g. in the volume adjustment or tuning tasks. Furthermore, some actions, such as storing radio stations as pre-sets, require a long ‘press-and-hold’ interaction, which exceeds the 0.2-second value of a keystroke. Consequently, it is important, when utilising the KLM, to consider the possible need for novel operators, or to consider how existing operators should be adapted in novel contexts, in order to improve absolute validity and error rates. A substantial benefit of the KLM lies in the ability to expand the method, adding new operators where appropriate. However, these operator values need

to be empirically determined. Efforts have been made in the research community to address this issue, particularly with navigation systems (see Manes et al., 1998, Green, 2003)

The KLM has been proven, here and elsewhere, to be applicable within the driving context. This study sought to further that work through an evaluation that remained faithful to the original technique as set out by Card et al. (1983), and by reporting its validity in predicting task times for in-car radios. Such systems are commonly cited as benchmark systems against which more novel devices are compared within the distraction literature. However, it should be stressed that the KLM cannot be pursued as a reliable measure of potential driver distraction in its current format. Static task time values are not considered robust, or sensitive, enough measures in isolation (see Chapter 3). The purpose of this study was not to investigate whether the traditional KLM can be used as a measure of distraction potential. Moreover, it provided an opportunity to confirm and extend research regarding its application in novel contexts (driving), and, to provide a working knowledge of the technique in order to inform future developments. The intention is now to adapt, or extend, the traditional KLM, in order to model human performance in other contexts – notably, the visually restricted conditions imposed on participants in occlusion trials. As discussed in Chapter 3, comparisons can be drawn between the 15-second rule and the occlusion technique; indeed, it has been proposed that the occlusion technique specifically addresses the shortcomings of the 15-second rule (Young et al., 2003, Green, 2003, Stevens et al., 2004). Therefore, a

logical next step is to use KLM techniques to model performance under the interrupted vision conditions of occlusion.

5.6 Summary

The study presented in this chapter sought to investigate the use of the KLM to predict task times associated with in-car interfaces. A range of tasks on two ICE systems was assessed empirically, and task time results compared with KLM predicted values. Strong correlations between the predicted and observed results suggest the KLM provides good relative validity within the context of in-vehicle systems. However, error rates were inconsistent between the two systems. Error rates were slightly weaker for the older system, and under-predicted performance times, whereas for the newer system predictions tended to over predict. This difference was considered a result of the more mechanical nature of interaction required by the older system, meaning that operator values were less realistic. Consequently, it is suggested that care be taken when using the KLM to make comparisons between two different interfaces, if the physical characteristics of input mechanisms are substantially different. Where this is the case, it may be necessary to develop new operators, or adapt existing ones, in order to account for these differences.

6 Study 3 – Feasibility study: extending the keystroke level model to assess the visual demand of in-vehicle information systems.

In Chapter 5, it was demonstrated that the KLM could be implemented in a vehicle context to predict IVIS task times (see also Chapter 3, Manes et al., 1998, Green, 1999b, Green, 2003). As such, the method has been proposed as a surrogate for the user trials required for assessing compliance with the 15-second rule (Green, 2003). However, there are concerns that the 15-second rule lacks the ability to distinguish between tasks with varying degrees of visual demand (Green, 2003, Young et al., 2003).

It has been proposed that the 15-second rule could be revised to include an occlusion component in order to satisfy these misgivings (Green, 2003). Chapter 4 investigated the validity of the occlusion technique, concluding that, whilst the method has merit, a significant drawback is its requirement to run performance based user trials. In this chapter, we examine the possibility of modelling performance of IVIS tasks under occlusion conditions at the keystroke level. If possible, an extended KLM may be used to predict the TSOT measure normally produced during occlusion user trials. Consequently, the second occlusion metric, R , could be determined as the ratio of the predicted TSOT (from the extended KLM) to static task time (from traditional KLM evaluations).

To investigate this novel implementation of the KLM, data from the occlusion study outlined in Chapter 4 is revisited. Observed results, focussing on two comparable tasks on two IVIS, are compared with the predicted values resulting from keystroke models of those tasks. Extended KLM models are based on traditional KLM teachings and logical assumptions for determining the placement of KLM operators in an occlusion framework.

6.1 Research questions

The questions addressed by this study are as follows:

1. Is it possible to adapt the traditional KLM to make predictions of the TSOT metric defined in the occlusion standard (ISO, 2007)?
2. What assumptions/rules need to be made regarding task performance under occluded conditions?
3. Do assessments of IVIS using an extended KLM draw the same conclusions as traditional, empirical occlusion studies?

6.2 Research contribution

In seeking to extend the KLM beyond the traditional full-vision paradigm, this study presents a unique contribution to research. To the author's knowledge, extensions of the KLM have previously focussed on generating new operators for the KLM, and thus have not furthered its use in such a novel manner.¹¹

¹¹ This sentiment was echoed by a reviewer of the CHI paper (Pettitt et al., 2007), who noted that such studies were not really 'extensions' since this flexibility was built into the original KLM concept.

6.3 Method

Occlusion trial data is the same as that gathered in the on-road evaluation study outlined in Chapter 4. The detail is repeated here for completeness, with further information provided where necessary in the context of this study.

6.3.1 Tasks

Tasks were selected as per the occlusion standard, which states that occlusion trials are only suitable for tasks in excess of five seconds when performed with full vision (ISO, 2007). Accordingly, two comparable tasks were selected for examination on each system: selecting a favourite destination, and adjusting sound configuration.

- **Navigation task:** for both systems, participants were required to navigate to the stored destination list from the home-screen. They were then required to select the requested destination from the list and confirm their choice. Selecting the required destination normally required participants to navigate through several pages of destinations.
- **Sound task:** participants were instructed to adjust various combinations of the volume, fade/balance and bass/treble settings of the systems. For System A, this required navigating to the sound configuration menu, and then adjusting the levels as required. For System B, the fade/balance and bass/treble settings appeared on separate menu screens.

6.3.2 Occlusion trials

Participants

Sixteen licensed drivers were selected to take part in the occlusion trials (mean age: 41.2; SD = 19.8; range: 22 – 68). There was an even gender split between participants. Participants had held their driving licence for an average of 22 years, and driving experience was recorded. Participants were paid for their time.

Equipment

The occlusion trials were conducted in a stationary vehicle. Participants were fully trained in the use of the two systems under investigation prior to trials. Of the two systems, one incorporated a commercially available, integrated navigation and entertainment system. The second was a prototype providing the same functionality as the commercially available system.

Occlusion was achieved using PLATO (portable liquid-crystal apparatus for tachistoscopic occlusion) goggles (Milgram, 1987, Milgram, 2005). These have been commonly used in other occlusion studies (see Stevens et al., 2004). The transparent/occluded cycle was controlled by software provided by the goggles' manufacturer.

Trials

The trials were conducted using the guidelines of the draft occlusion standard as available at the time (ISO, 2004a). In a repeated measures design, each participant completed two trials, one for each system, with a delay of around two to four weeks between trials. To avoid order effects, the participants were

split into two groups, with one group completing the trials in the order System A then System B, and the other in the order System B followed by System A. Participants were trained in the tasks and given five practice trials in order to familiarise themselves with the task/system. An assessment was made of their ability to successfully continue with the trial at this point, since both the occlusion technique and the KLM expect skilled levels of performance. As per the occlusion standard, each participant completed five variations of each task with full vision, and five variations under occlusion conditions, during trials. The ordering of full vision and occlusion conditions was counterbalanced to minimise learning effects. Task timing began once the investigator completed the task instruction and ended when the participant had successfully completed the task.

Results

Some technical issues encountered during the occlusion trials reduced the initial data set of 16 participants. For static task times, results were obtained for 14 participants in the navigation task and 15 participants for the sound task on each system. In the occlusion trials data was collected for 15 participants for each task on the commercial system, whilst data for only 12 participants in the navigation task and 11 in the sound task, was collected from the prototype.

The occlusion standard requires the reporting of four statistics for each occlusion measure (TSOT and R). The results are summarised in Table 6.1. No significant differences were detected between the two systems for the navigation task for either the TSOT or R-values. For the sound task, a

significant difference was detected between the systems for each measure. For TSOT, System A's value was found to be significantly smaller than System B ($t(23)=-3.48$; $p<0.01$), whilst for the R ratio, System A's value was found to be significantly larger than System B ($t(24)=3.08$; $p<0.01$).

Table 6.1 Summary of occlusion statistics from data gathered during user trials

		Navigation Task		Sound Task	
		System A	System B	System A	System B
TSOT	Mean	9.3	9.4	9.1	13.9
	Median	9.1	9.1	8.0	14.0
	Standard Deviation	2.7	1.8	3.7	3.3
	85 th Percentile	12.3	11.1	12.9	16.6
R	Mean	0.94	0.84	0.69	0.51
	Median	1.11	0.87	0.66	0.56
	Standard Deviation	0.33	0.20	0.17	0.12
	85 th Percentile	1.23	1.04	0.83	0.61

6.3.3 KLM models

Predicting static task time

Keystroke level models (KLM) were developed for each task on each system. Initially, tasks were described in systematic prose. Next, flow charts were developed. Finally, a sequence of operators necessary for performing the task was written. The operator values used were generally taken from those originally laid out by Card et al. (1983). However, the 'reach far' operator (Rf) (Green, 2003) was substituted in place of a homing operator (H), as a more accurate prediction of the time necessary to move the hand from the steering wheel to the system controls. The nature of the KLM is such that new operators are created or existing operators are adapted as novel interfaces are designed. A selection of operators that featured most prominently in this experiment is provided in Table 6.2.

The nature of the occlusion trials was such that participants completed several variations of each task. As a consequence there were slight differences in task structure and length between variations in the occlusion trials (i.e. longer/shorter street names, town names, and so on). It should be noted that this variation was not related to the participant performance. It was therefore necessary to produce several KLMs for the sound task on both systems, and the navigation task on System A, in order to make accurate comparisons between the two methods.

Table 6.2 Selection of operators used in KLM models

Operator	Action	Time (seconds)
Rf (Green, 2003)	Move hand between steering wheel and in-vehicle device	0.45
M (Card et al., 1983)	Mentally prepare for an action	1.35
K (Card et al., 1983)	Press a key	0.2

Predicting TSOT

Once satisfactory KLMs were completed for each task/variation, predictions of static task time could be made. The next stage was to attempt to predict the TSOT values. The operator sequences were assessed against a time line of the vision/no-vision sequence as experienced by participants during occlusion trials. Figure 6.1 illustrates the principle of extending KLM models against the occlusion cycle.

Assessing the static operator sequence against a timeline of vision/non-vision periods requires the modeller to make decisions concerning the effect of visual occlusion on task performance. In this respect, it was necessary to formulate a set of guidelines, similar in fashion to the heuristics provided by Card et al. (1983), for the placement of mental operators within a KLM. These were formed from previous experience of occlusion trials (see Chapter 4) and

knowledge of the occlusion literature (see Chapter 3). This combined experience suggests that the fundamental purpose of these guidelines is to guide a modeller in determining those steps in a task that may afford blind operation. Accordingly, the following logical assumptions were made:

- **Assumption 1:** during 1.5 second periods of vision the operator sequence progresses without interruption;
- **Assumption 2:** an operator that begins in a period of vision can continue into a two second occluded period providing it is not specifically associated with vision, for example reading information from a display;
- **Assumption 3:** an operator can only begin in an occluded period when vision is not required at any point in its duration, for example a keystroke where the finger is already placed on the control in question.

Using these assumptions, static operator sequences were adapted to model performance under occlusion conditions. The number of vision periods could now be counted, the result multiplied by 1.5, and any partial vision periods at the task conclusion taken into account, to give the predicted TSOT. As with occlusion trials, the predicted value of R is given by dividing the value of the TSOT by the predicted static task time.

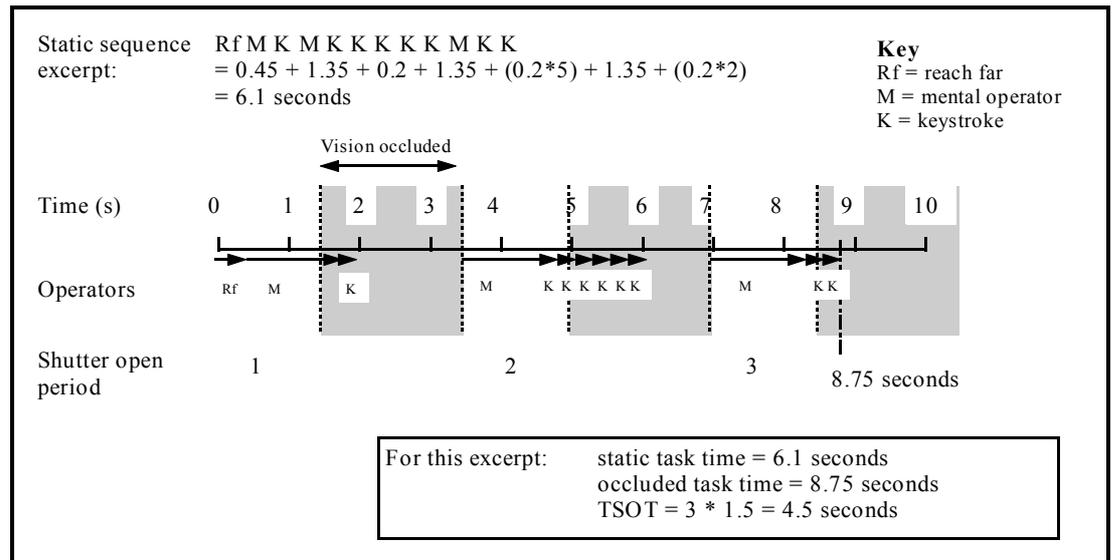


Figure 6.1 Example of modelling occluded performance

Prediction results

The values for TSOT and R predicted by the extended KLM method are summarised in Table 6.3.

Table 6.3 Summary of extended KLM predictions for TSOT and R

	Navigation Task		Sound Task	
	System A	System B	System A	System B
TSOT	9.3	10.5	9	11.7
R	0.88	0.93	0.77	0.54

6.4 Comparison between methods

In comparing predicted values with observed results, one approach may be to use occlusion measures to judge which system's results are most desirable. Put simply, is System A better/worse than System B? This is in keeping with the spirit of the occlusion technique, which is considered a cost-effective method for assessing systems during development (Stevens et al., 2004). Furthermore, the lack of clear indications regarding acceptable values for TSOT and R within the draft standard suggests such an approach is valid. As such, figures 6.2 and 6.3 illustrate the comparisons between the two systems investigated in

this study, showing occlusion measures (mean, median, 85th percentile) and the extended KLM predictions.

As noted earlier, significant differences between systems were only found for the sound task. For TSOT, both the observed occlusion measures and the predicted KLM values favour System A over System B. However, for the R-ratio the order is reversed. For the navigation task, there is discrepancy within the occlusion measures for TSOT. Whilst the predicted measures suggest that System A is more desirable than System B, no clear preference is indicated in the observed results.

An alternative approach is to evaluate the system results against the limit for TSOT of 15 seconds, as set out by the AAM (2003), and an R value smaller than one, a limit implied by the occlusion standard (ISO, 2007). Under these conditions, it may be concluded that both tasks on each system are acceptable according to predicted measures and mean observed values.

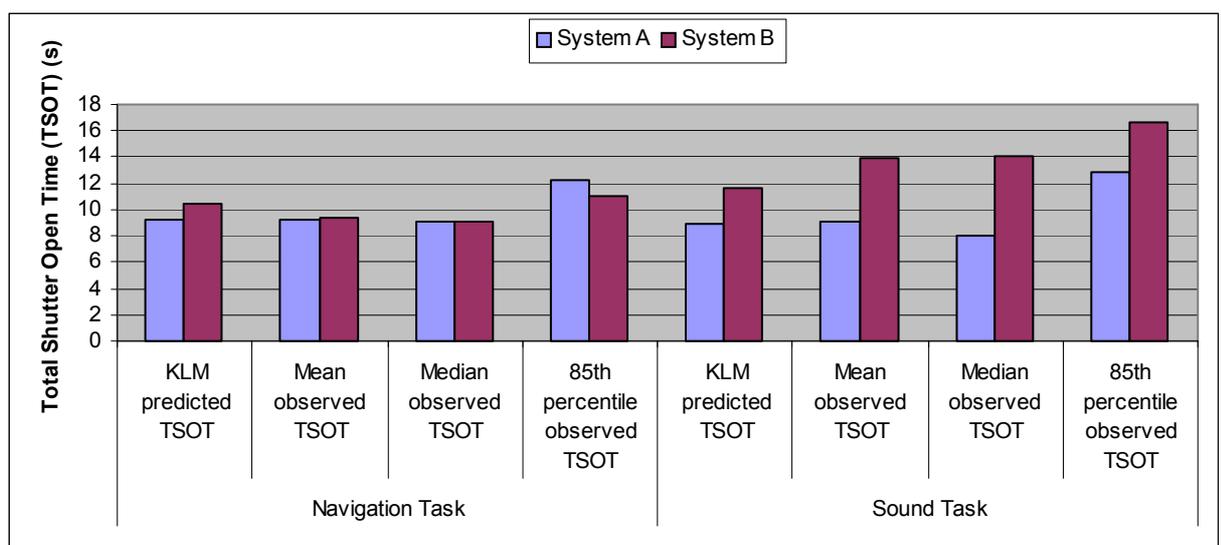


Figure 6.2 Comparison of predicted and observed TSOT values

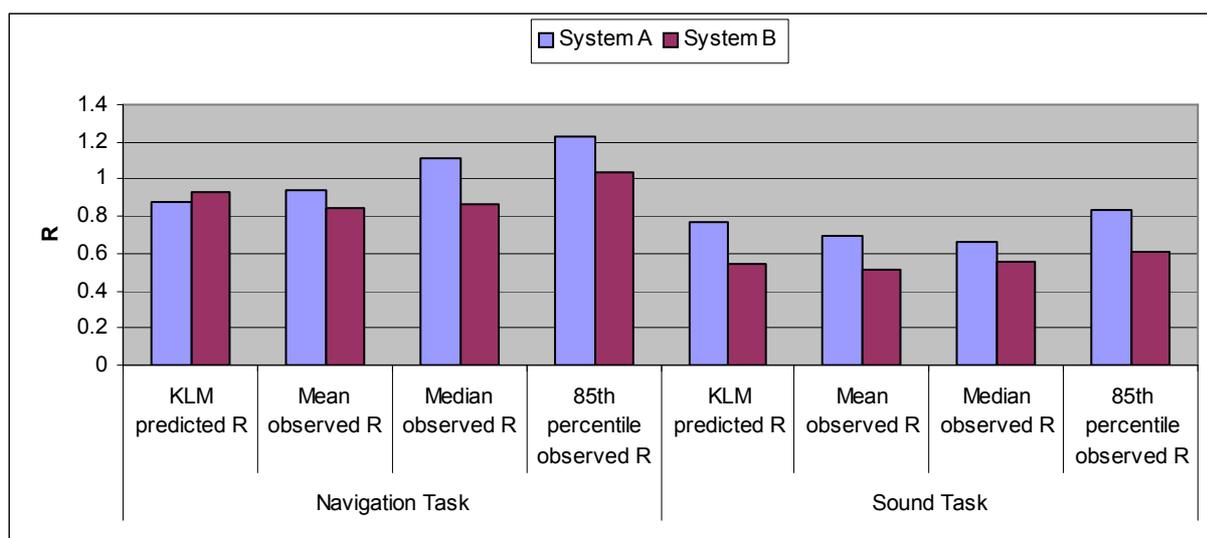


Figure 6.3 Comparison of predicted and observed R-values

6.4.1 Error rates

KLM studies commonly report absolute prediction error as a measure of the technique's validity (e.g. Luo and John, 2005). For this study, the prediction error rates for the TSOT and R values are presented in Table 6.4. The results show low error rates for TSOT prediction for System A, whilst for System B results are less strong. However, all values are in keeping with the expectation of predictions within 20% of skilled execution time (John et al., 2004b).

Table 6.4 Prediction error rates (%) for TSOT and R values (calculated as follows: difference between 100 and predicted task time as a percentage of actual task time, consistent with John et al., (2004a))

	Navigation Task		Sound Task	
	System A	System B	System A	System B
TSOT	0	11.7	- 1.10	- 15.8
R	- 6.4	10.7	11.6	5.88

6.5 Discussion

The results presented in this chapter are encouraging, suggesting there is potential for occlusion measures to be predicted without the necessity of running user trials. Consequently, assessments of visual demand can be made during earlier stages of design, as predictive methods do not require the robust

prototypes necessary for user trials (Green, 2003, Salvucci, 2005). Modelling techniques, such as that discussed here, could potentially be applied once simple designs have been drawn and task procedures formalised. This, in turn, makes it easier for designers to make significant adjustments to proposed systems.

Two approaches to the analysis were considered. The first was to view each technique as a method for comparing two interface designs for identical tasks, and in this way establishes occlusion measures as part of a formative evaluation process. In this fashion, the techniques may be used to make design decisions – put simply, the system with the more preferential result would be selected for further development. In the user trials, a significant difference between the systems was only detected for one of the tasks, the sound task. For the measure of TSOT, System A was preferred over System B. For the R-metric, this decision was reversed, with System B having the more desirable value. In both cases, the predicted KLM values would have led to the same design decision as the user trials. For the second task, the navigation task, the predicted KLM anticipated that System A would be preferred over System B; however, the predicted difference for both measures, TSOT and R, was small. As stated, the results of the occlusion trials failed to detect a significant difference between the systems, therefore no firm design decision can be made. It could be expected that a larger sample of participants would demonstrate a difference between the systems in line with the predictions.

In a second approach, the results from both the model predictions and the user trials were assessed against limits for TSOT and R proposed in the literature. Following this approach, measures are used as part of a summative evaluation, against pass/fail criteria. A task which exceeds agreed limits may be locked-out from driver engagement whilst a vehicle is in motion, or removed from the system altogether. Here, only very cautious conclusions can be drawn using this approach, as both tasks on each system were considered acceptable against these criteria in both the predicted and observed results. Of some concern may be that, for the sound task on System B, the mean observed TSOT was approaching the limit of 15 seconds, at just under 14 seconds, whilst the predicted value was less than 12 seconds. One suggestion might be that limits for predicted values be set more conservatively, however this may be inappropriate, given that the prediction for System A is very similar to the observed result. Furthermore, the error rates show that the predicted TSOT value for the sound task on System B was closer to 20%, yet for System A, they were less than 2%. Consequently, it may be the case that, for System B, participants' performance in occlusion trials was more variable than the model accounted for; for example, if participants employed different methods for completing the task. Alternatively, it may suggest an intrinsic error in the task models. System B was a prototype system, and presented many difficulties throughout the trials. Specifically, the prototype was not robust enough to cope with the demands placed on it by user trials. This highlights the need for a well-developed prototype in user trials where human performance measures are to be taken. Because of the system's failings, performance may not have been at the expert, routine level anticipated by the KLM.

6.6 Summary

This chapter presented a study assessing the feasibility of extending the KLM to consider human performance in occlusion trials. Occlusion user-trials data from the study outlined in Chapter 4 was revisited and compared with results predicted by an extended KLM approach. A series of three logical assumptions were developed, which assist a modeller in the placement of KLM operators against a timeline of the interrupted vision sequence of the occlusion paradigm. The results demonstrate there is merit in the approach, with the predicted measures leading to the same design decisions as the observed results. However, the user-trial data considers only two in-vehicle tasks, and there were concerns over user-performance with one of the systems. Therefore, a further validity study that considers a wider range of tasks is recommended.

7 Study 4 – Validation of the extended KLM

The previous chapter investigated the possibility of extending the KLM to predict the metrics normally measured through performance based user trials with the occlusion technique. This approach was inspired by the use of the KLM in assessing compliance with the 15-second rule, and other studies demonstrating that the technique could reliably be used to predict static times for IVIS tasks. Modelling techniques, such as the KLM, enable quantitative assessments of task performance to be made much earlier in the design cycle than is possible with user-trial based methods. If the metrics generated in occlusion trials can be predicted through keystroke-level modelling, it will be possible to realise the benefits of the technique whilst reducing the associated costs.

Study 3 demonstrated that there was merit in such an approach. Logical assumptions were made to guide modellers in the placement of KLM operators within the interrupted vision sequence experienced by participants in occlusion trials. However, it was concluded that in order to present the new technique as valid, a further study was necessary. In this case, it would be useful to examine a wider range of tasks, on systems with a reliable level of performance. The study outlined in this chapter presents results from a new occlusion study focussing on three comparable IVIS tasks on two systems, and compares the results with models of those same tasks.

7.1 Research questions

This study was designed to address the following questions:

1. Is the extended KLM a valid technique for predicting the measures associated with the occlusion technique, when assessed against a wider range of tasks on reliable systems?
2. Were the logical assumptions made in the previous study theoretically sound?

7.2 Research contribution

The study reported in this chapter advances the work of the previous chapter in validating the extended KLM and its associated logical assumptions made regarding human performance under occlusion conditions.

7.3 Method

To evaluate the extended KLM technique, three tasks on two IVIS were assessed in occlusion trials. Following this, KLM predictions of static task time, TSOT and R were made, and the results compared with the user trial data. The study was piloted prior to trials and ethically approved by the School of Mechanical, Materials and Manufacturing Engineering at the University of Nottingham.

7.3.1 Occlusion

Participants

Twelve licensed drivers were selected to take part in the occlusion trials, six male and six female. Participants ranged in age from 20 to 55+ (age not disclosed), had held their driving license for an average of 15 years (SD=13.6; range: 3 to 41), and drove on average approximately 8,000 miles in the last

year (SD= 6 500; range 1 000 to 20 000). Of the participants, ten had no prior experience with in-car navigation systems, and of the two that had, neither had experience with the specific systems investigated in the trials.

Equipment

Two nomadic satellite navigation systems were used in the trials, a Navman iCN 520 (System A) and a TomTom Go 300 (System B) (see Figure 7.1). Trials took place in a stationary, modern, family sized car. Occlusion was achieved with Plato occlusion goggles (Milgram, 1987, Milgram, 2005). These goggles are the generally preferred tool used in occlusion studies in the literature (see Stevens et al., 2004).



Figure 7.1 System A (left) and System B (right)

Tasks

Three common, comparable tasks on each system were selected for investigation. These were entering a destination; selecting a destination using the Point of Interest (POI) function; and, adjusting the volume.

- **Destination entry (DE):** users were required to navigate a menu system in order to reach data entry screens. Each system structured the data entry by town, street and house/building number. System A had extra menu stages in comparison with System B. For example, after

entering the name of the town and confirming, with System B users are immediately taken to the street input stage. With System A, an intermediary step is taken, where users confirm they wish to progress to the street entry stage. The data-entry keyboards on the touch-screens for both systems followed an ABC layout as opposed to the QWERTY common with standard desktop computer keyboards.

- **POI search:** users were required to access a stored list of destinations ordered by category (e.g. airports, train stations, shopping centres). For both systems, POI lists were accessed through the menu system, however, similar to the destination entry task, System A required participants to take more steps to reach the lists. Furthermore, whilst System A ordered POI items alphabetically, System B ordered items based on proximity.
- **Volume change (VC):** participants were asked to adjust the volume to maximum, minimum, high, low or medium values. For System A, this required several steps through a menu system to get to the volume screen, where the volume was adjusted by pressing a point on a scale. For System B, users were required to press their finger in the lower left portion of the default screen; upon doing so, a volume bar was displayed and the desired level selected.

Training

Participants were trained in the use of the navigation systems. After an introduction to the system and its controls, participants were guided through the tasks under investigation. Following this, participants completed five

practice trials of each task. During training participants were also made familiar with the occlusion goggles, and a minimum of two practice trials were completed under occlusion conditions. This was in accordance with the terms of the draft occlusion standard as available at the time (ISO, 2005).

Trials

In a repeated measures design, each participant completed two trials, one for each system. The participant group was split, and exposure to the systems counterbalanced. Each participant completed five variations of each task with full vision, and five under occlusion conditions. The ordering of full vision and occlusion was counterbalanced. In occlusion conditions, participants' vision was interrupted on a cycle of 1.5 seconds vision followed by 2 seconds occluded vision. This was in accordance with the version of the international standard available at the time of the trials (ISO, 2005). The full standard suggests a cycle of 1.5 seconds vision followed by 1.5 seconds occlusion (ISO, 2007). Research suggests that a broad range of shutter closed periods are acceptable (Stevens et al., 2004), and this is not felt to be a critical change – unlikely to have any impact on the results reported here.¹²

Results

The results of the occlusion trials are presented in Table 7.1. The ISO standard for occlusion requires that the metric TSOT be reported from occlusion trials, with the option to report R (ISO, 2007). In each case, studies should report mean, median, standard deviation and 85th percentile values for each metric.

¹² Stevens, personal communication

However, the standard provides no indication as to how these results should be interpreted. Median and 85th percentile values are not necessary for the later comparisons with KLM predictions, and are therefore not presented here.

Paired two sample t-tests were performed on the data. For TSOT, highly significant differences were detected between the systems for all tasks (DE: $t(11)=10.65$, $p<0.001$; POI: $t(11)=6.98$, $p<0.001$; VC: $t(11)=9.58$, $p<0.001$). For the R ratio, a significant difference was only detected between the systems for the DE task ($t(11)=-6.16$; $p<0.001$).

Table 7.1 Occlusion results showing means and standard deviations in parenthesis

	System A			System B		
	DE	POI	VC	DE	POI	VC
Static time	52.75	16.32	8.74	20.58	9.47	3.26
– seconds	(7.16)	(2.10)	(1.93)	(5.12)	(1.64)	(1.06)
TSOT –	38.73	11.90	7.69	20.07	7.09	2.28
seconds	(9.75)	(2.20)	(1.63)	(6.20)	(1.45)	(0.90)
R	0.74	0.74	0.90	0.98	0.76	0.74
	(0.16)	(0.14)	(0.19)	(0.18)	(0.15)	(0.31)

7.3.2 Extended KLM

Following collection of the empirical, occlusion data, the tasks as undertaken by the participants were modelled in order to make KLM predictions of static task times and TSOT. With values for these two metrics, a predicted value for R could also be calculated.

Predicting static task time: operator values and Fitts' law

For each variation of the three tasks completed by participants in static (full vision) conditions, a KLM was drawn up. Tasks were decomposed into their

basic operators, with mental operators being inserted according to heuristics laid out by the KLM's authors (Card et al., 1983). This provided the basis for predicting static task time. Operator values for the mental and keying operators were taken from the original KLM (Card et al., 1983). For movements between the steering wheel and the in-vehicle device, a 'reach far' (Rf) operator was used (see Green, 2003).

The nature of the tasks and systems involved in this experiment meant that when keying information into the systems, participants would likely use just one digit (e.g. their index finger), unlike desktop computing tasks where a KLM may assume touch-typing performance. Therefore, it is necessary to consider the movement time between keys when building KLMs. Measuring human performance empirically can approximate such times, which is how the original KLM operator times were ascribed, or, alternatively, they can be predicted using a formula known as Fitts' law (Card et al., 1983, Raskin, 2000). Estimating the time to travel to a target as a logarithmic function of the distance to a target and the target width, Fitts' law has been utilised to predict pointing, or cursor, movements (P) (Raskin, 2000). In this study, Fitts' law was used to predict the time taken to move the input finger between buttons on the in-car devices (H).

For a full explanation of the development of a KLM, and the general theory surrounding the method, including the heuristics for the placement of mental operators, and the use of Fitts' law, readers are directed to the method's authors, Card et al. (1983). In addition, Raskin (2000) introduces the technique

in an accessible format. A discussion of operators specific to the driving context can be found in Green (2003). For the purposes of this study, traditional operator values were generally used. However, the reach far (Rf) operator, for movement between the steering wheel and the IVIS was used, presented in Green (2003), and Fitts' law was used to estimate times for movements between controls on the IVIS.¹³

Predicting TSOT: extended KLM models

In order to make predictions of Total Shutter Open Time (TSOT) – which is the amount of visual time in occlusion trials – KLM models were corresponded with a timeline of the vision/non-vision cycle as experienced under occlusion conditions. This requires the modeller to make decisions as to what degree the occlusion process disrupts an operator sequence. A series of logical assumptions were made as part of a feasibility study (see Chapter 6):

- **Assumption 1:** During 1.5-second periods of vision the operator sequence can progress without interruption;
- **Assumption 2:** An operator that begins in a period of vision can continue into a two second occluded period providing it is not specifically associated with vision, for example, reading information from a display;
- **Assumption 3:** An operator can only begin in an occluded period when vision is not required at any point in its duration, for example a keystroke where the finger is already placed on the control in question.

¹³ Note: this was not considered necessary in the previous study (Chapter 6) since the interaction mechanism on those systems allowed users to complete tasks without moving significantly between controls.

The development of an extended KLM has two stages.

- **Stage 1:** a traditional KLM for the task under investigation is produced, following guidelines presented in Card et al. (1983), Raskin (2000), and Green (2003).
- **Stage 2:** the KLM is then assessed against the vision/no-vision timeline as would be experienced by participants in user trials with the occlusion technique. The three logical assumptions outlined above are used to determine at which points the non-vision periods would likely result in a pause in the task performance.

Example of extended KLM

To illustrate the principle, the following demonstrates how a task is assessed under the extended KLM technique. The example uses the first section of the destination entry task, from the start of the task to the point where the participant would begin to enter alphanumeric data, for System A. The example is illustrated in Figure 7.1.

Stage 1: traditional KLM operator sequence developed

For this section of the task, the user is required to move their hand to the system and press the ‘Go to’ button on the menu screen. On the following menu screen, they must press the ‘Address or POI’ button. In the final stage, they are required to press the ‘Find area’ button. Following the KLM guidelines (Card et al., 1983, Raskin, 2000, Green, 2003), the following operator sequence is produced:

1. Move hand to system (Rf, 0.45 seconds)
2. Mentally prepare (M, 1.35 seconds)
3. Press 'Go to' button (K, 0.2 seconds)
4. Mentally prepare (M, 1.35 seconds)
5. Move to 'Address or POI' button (H, 0.62 seconds)
6. Press 'Address or POI' button (K, 0.2 seconds)
7. Mentally prepare (M, 1.35 seconds)
8. Move to 'Find area' (H, 0.62 seconds)
9. Press 'Find area' (K, 0.2 seconds)

This sequence of 9 operators predicts a static task time of 6.34 seconds.

Stage 2: development of extended KLM

The sequence is now assessed considering the vision/no-vision intervals of 1.5 seconds (vision) and 2 seconds (no-vision) as experienced by participants during occlusion trials, making decisions based on the three assumptions above:

1. Operator 1 (Rf) takes place as normal, according to assumption 1.
Elapsed time 0.45 seconds.
2. Operator 2 (M) begins following operator 1, crossing into a no-vision period after 1.05 seconds, following assumption 2. *Elapsed time 1.85 seconds.*
3. Operator 3 (K) cannot begin until the next period of vision, at 3.5 seconds, following assumption 3. *Elapsed time is 3.7 seconds.*
4. Operator 4 (M) crosses into a no-vision period after 1.3 seconds, following assumption 2. *Elapsed time 5.05 seconds.*

5. Operator 5 (H) begins at start of next vision period, at 7 seconds elapsed time, according to assumption 1. *Elapsed time, 7.62 seconds.*
6. Operator 6 (K) is completed according to assumption 1. *Elapsed time 7.82 seconds.*
7. Operator 7 (M) crosses into a no-vision period 0.68 seconds into its duration, following assumption 2. *Elapsed time, 9.17 seconds.*
8. Operator 8 (H) begins at start of next vision period, at 10.5 seconds, according to assumption 1. *Elapsed time, 11.12 seconds.*
9. Operator 9 (K) follows the previous operator uninterrupted, according to assumption 1. *Elapsed time, 11.32 seconds.*

This extended KLM predicts a total task time under occluded conditions of 11.32 seconds, with a TSOT value of 5.32 seconds. Given the static task time prediction (6.34 s) and the TSOT prediction (5.32 s), we can now predict an R-value for this model of $5.32/6.34 = 0.84$.

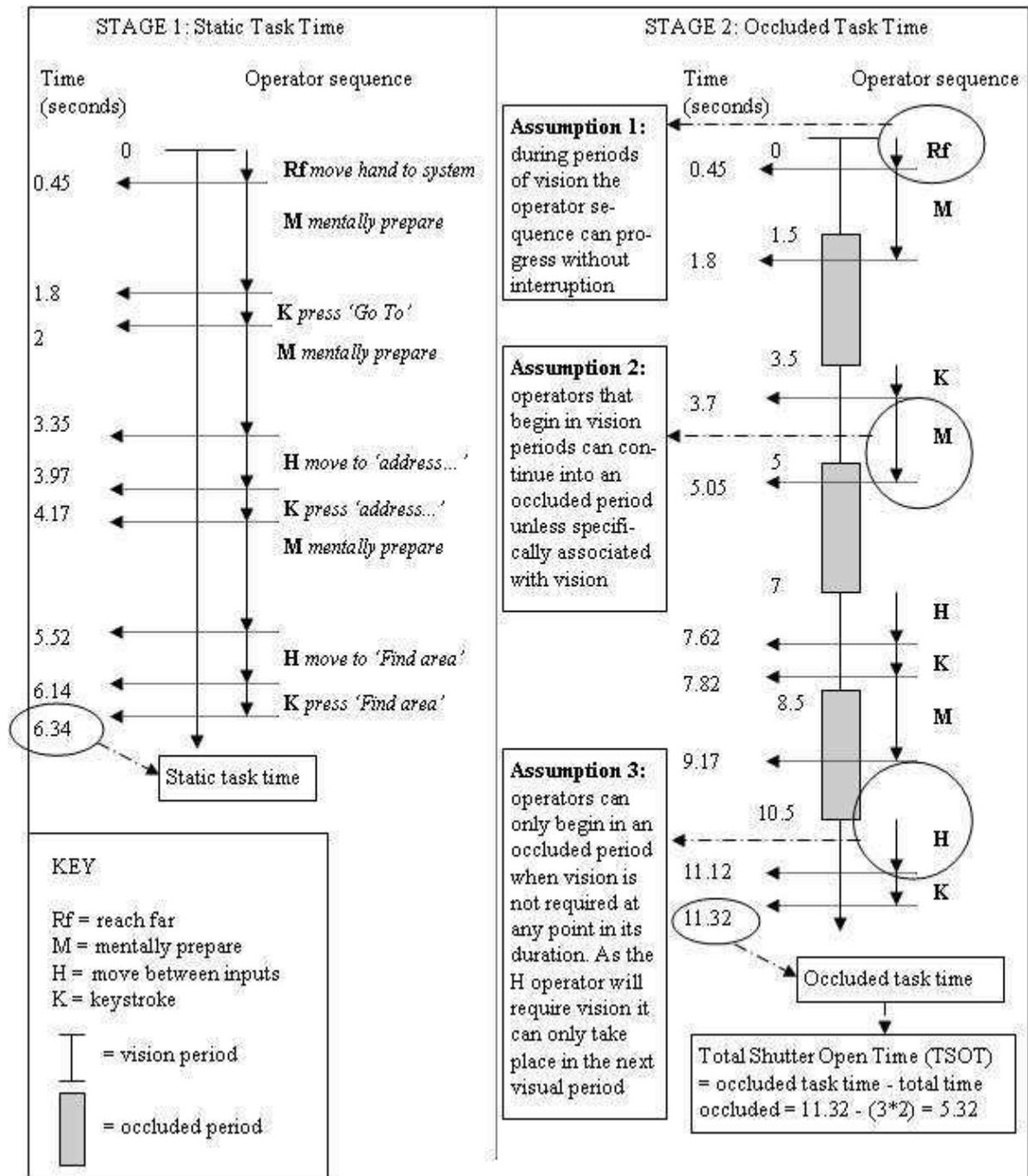


Figure 7.2 KLM operator timelines illustrating first portion of the destination entry task for System A without occlusion (left) and with occlusion (right)

KLM Predictions

The predicted values for the static task time, TSOT and R-ratio results of the tasks as undertaken by participants in the occlusion trials are presented in Table 7.2.

Table 7.2 KLM predictions of static task time, TSOT and R

	System A			System B		
	DE	POI	VC	DE	POI	VC
Static time	41.12	16.59	9.61	18.46	10.61	3.12
TSOT	27.82	12.91	7.53	17.68	8.44	2.53
R	0.68	0.78	0.79	0.96	0.80	0.81

7.4 Comparison between predicted and observed results

Error rates for the KLM predictions compared with the mean results of the occlusion trials are presented in Table 7.3. Correlations between predicted and observed results for TSOT and static task time were strong. For static task time, a correlation coefficient of 0.98 was recorded, for TSOT of 0.93.

Table 7.3 Error rates (%) of KLM predictions (calculated based on predicted task time as a percentage of actual task time, consistent with John et al., (2004a))

	System A			System B			Mean (+/-)
	DE	POI	VC	DE	POI	VC	
Static task time	-22.1	1.65	13.2	-10.3	12.0	-4.21	10.6
TSOT	-28.2	8.49	-2.03	-11.9	19.0	11.1	13.5
R	-7.85	6.72	-12.1	-1.79	4.67	9.46	7.10

7.5 Discussion

The results of this study are encouraging. By using an extended version of the KLM, the author sought to predict the measures of TSOT and R obtained from an occlusion trial. The KLM has previously been shown to make reliable predictions of performance times by expert users in routine tasks (Card et al., 1983, Raskin, 2000). Other research has demonstrated that the technique can be implemented in situations other than desktop computing, for which it was originally developed (e.g. Manes et al., 1998, Stanton and Young, 2003; Myung, 2004, Luo and John, 2005). The main advantage of the technique lies in its ability to model human performance, therefore reducing the reliance on

user trials, and providing useful information in early design stages. The occlusion technique is a user-trial based method for assessing the visual demand of an IVIS. Numerous studies have concluded that there is merit in the technique (see Chapter 3). However, a perceived limitation of the technique's effective use by IVIS designers is the necessity to evaluate through performance-based user trials. In the study reported here, the author sought to address this concern.

In comparing the predicted values with the observed results, correlations were strong and error rates were generally low. When applied correctly, practitioners can expect KLM predictions to be well within 20% of skilled execution time (John et al., 2004b). Of particular note, the results suggest that errors in the predictions lie within the traditional keystroke models, and are not overly compounded by the process of extension; for example, average error for static task time was 10.6% and 13.5% for TSOT. This suggests that the assumptions made regarding task performance under occluded conditions are valid.

The error rate for the destination entry task on System A was particularly high. However, as with the other tasks, the issue appears to lie with the traditional KLM aspects rather than the extension process. During user trials, participants regularly deviated from the prescribed task method, and made occasional errors. These factors are likely to have had an impact on the accuracy of the predictions, which are based on the assumption of expert, error free performance of a particular route through the task.

According to AAM and ISO guidelines on acceptable values for TSOT and R, the predictions accept/reject the tasks in the same fashion as the observed results (AAM, 2003, ISO, 2007). Taking the AAM suggested limit of 15 seconds for TSOT, both the predictions and the observed data would pass the POI and volume change tasks for each system; both measures would fail the destination task.

Arguably, the most troublesome aspect of applying the KLM lies in the placement of mental operators. A set of heuristics has been developed to aid practitioners (Card et al., 1983), although there is still scope for variability. If a practitioner is over-zealous with the placement of Ms, the resulting prediction may be an overestimate. Alternatively, the prediction may assume a much more automated level of performance, resulting in an underestimate. If difficulties are experienced, a solution may be to base a prediction on an average of two KLMs, one overestimation and one underestimation.

7.6 Summary

The study outlined in this chapter supports the work of the previous chapter, suggesting there is considerable merit in the extended KLM technique. Three tasks on two systems were assessed in an occlusion study and metrics compared with predictions made through an extended KLM approach. Error rates were consistently low and correlations between predicted and observed measures were strong. The results suggest that the three logical assumptions regarding task performance under occluded conditions are valid.

8 Case Study – Reliability of the Extended KLM

For an evaluation method to be usable and practical, validity in itself is not sufficient; the method must also be reliable. By reliable, what is meant is that the method can be implemented independently and yield similar results. With user-trial based methods, one way to improve reliability is through the standardisation process. Standardisation presents experimenters with guidelines that specify how participants should be selected and trained, how trials should be conducted, and how measurements should be taken. For example, the occlusion standard provides a framework through which the method is applied (ISO, 2007). The results from two independent studies that both follow this framework can therefore be assumed comparable, although there is a lack of research to confirm this. However, when user-trial methods vary in application then reliability is likely to be questionable. This problem may be encountered when seeking to compare the results of a variety of simulator, or road based, trials, since there is less standardisation across study procedure.

There is sufficient evidence to suggest that the traditional KLM is a reliable method. Since its inception in the 1980s, a large body of research has shown valid applications of the method, even transporting it beyond the original desktop computing paradigm for which it was conceived (see Chapter 3 for a fuller discussion). Particularly encouraging is the study by Stanton and Young (2003), which demonstrated that, of a variety of ergonomics methods, the KLM was the most easily learnt and applied by novice users, since a criticism of the

technique has been that, conversely, it is difficult to learn (John, 2003, Dix et al., 2003). Fundamental to the reliability of the method is the set of heuristics provided by Card et al. (1983), which guide modellers in the placement of mental operators. The rigorous application of these rules should lead to accurate predictions of task time values (John et al., 2004b).

The extended KLM adds a further layer of subjectivity to the original method, and it is here that assessments of reliability are necessary. The new technique involves modelling, at the keystroke level, of performance under the interrupted vision conditions of the occlusion technique. Essentially, this translates to making decisions concerning the performance of operators in non-vision periods. Similar to the heuristics for the placement of mental operators, the extended KLM provides a set of assumptions regarding occlusion performance to guide modellers in making these decisions. Studies 3 and 4 tested the validity of these assumptions with promising results. In order to test the reliability of these assumptions it is necessary to evaluate whether an independent modeller applies them consistently.

As part of a wider evaluation of the occlusion technique, commissioned by the UK Department for Transport, with the Transport Research Laboratory (TRL) as the leading investigators, an extended KLM evaluation of four IVIS tasks was performed. The predicted results and those obtained from standard user-trials are discussed below in terms of the reliability of the extended KLM.

8.1 Research questions

The research questions addressed by this chapter are as follows:

1. Can the extended KLM technique be reliably applied by an independent modeller, to make valid predictions of occlusion trial results?
2. Does a reliability study provide any insight into the relative economy of the extended KLM?
3. Do the logical assumptions of the extended KLM require any further development in light of the modeller's experiences?

8.2 Research contribution

The case study presented in this chapter demonstrates the validity of the extended KLM when applied independently and compared with exhaustive occlusion trials, thus allowing for some inference of the technique's reliability. Further, two tasks in this trial deviate in nature from those previously considered, and are likely to be more visually demanding, therefore providing an opportunity to further assess the extended KLM's validity. Next, the study allows for some evaluation to be made of the technique's economy, through a comparison of its relative costs in comparison with the occlusion trials. Finally, the use of an independent modeller provides an opportunity to assess comprehension of the extended KLM heuristics.

8.3 Tasks

The four IVIS tasks investigated were POI search, destination entry, and two variations of a scrolling text task, simple and complex. For the POI search, the procedure followed was broadly similar to the task as described in Chapter 7. The destination entry task required participants to input the first four letters of

a city name, select the city from a list, select the required street from a list and then enter the house number. This is in contrast to the longer method used in the previous study, where users were asked to enter entire destinations manually. In the scrolling text tasks, participants were required to identify three letter company codes and read out associated stock prices. These bear similarities to tasks implemented in Noy et al. (2004), where R values in excess of 1 were observed. In the simple scrolling text tasks, company codes and stock codes were presented in a single, continuously scrolling column. For the complex variation, three columns were presented simultaneously (Horberry et al., 2007b).

8.4 Occlusion trials

Occlusion trials were conducted at TRL in Berkshire (Horberry et al., 2007b).

8.4.1 Participants

Sixty participants took place in the trials. Ten participants were selected from each of six age categories (17-26; 27-36; 37-46; 47-56; 57-66; 67-76). There was an even gender balance across all participants, and all held full, valid driving licences.

8.4.2 Equipment

Tasks were conducted on a PDA device loaded with TomTom™ navigation software. The scrolling text task used a simple program developed for the trials. Interaction took place through a touch-screen mechanism. Occlusion was achieved using PLATO goggles. Trials took place in a stationary

Vauxhall Astra, with participants in the driving seat and the in-vehicle device positioned in the centre-console area of the vehicle.

8.4.3 Trial procedure

Trials followed the draft occlusion standard as available at the time (ISO, 2006). In contrast to the earlier studies reported within this research, the occlusion interval in this trial was set at 1.5 seconds as opposed to two seconds. Participants were trained in task procedures from written instructions relayed verbally, and given five practice attempts on each task. They then performed five static and five occluded trials for each task. Each of the task instances were distinct but of comparable difficulty. The ordering of static and occluded trials, and the order of the tasks, was randomised to minimise order effects.

8.4.4 Results

The results of the occlusion trials are presented in Table 8.1. The table presents mean values for TSOT and the mean of ratios for R. This is in accordance with the draft ISO standard. For each of the sixty participants, an R-value is calculated based on the means of their five static and five TSOT values. The mean of these ratios is then calculated.

Table 8.1 Mean TSOT and R observed values (with Standard Deviations)

Task	TSOT	R
POI Search	7.63 (2.75)	0.88 (0.45)
Destination entry	11.77 (3.16)	0.81 (0.17)
Scrolling text (short)	10.94 (5.27)	0.78 (0.38)
Scrolling text (long)	11.62 (5.69)	0.87 (0.39)

8.5 Extended KLM predictions

An extended KLM evaluation of the tasks outlined above was conducted at the University of Nottingham independently of the user trials (Horberry et al., 2007a).

8.5.1 Modeller

Extended KLM predictions were made by an academic member of staff from the School of Computer Science and Information Technology at the University of Nottingham. The modeller had prior experience of the traditional KLM, through his position as a lecturer on Human Factors methods and user-centred design. He was also familiar with the occlusion technique through a number of research projects, and had a good understanding of the extended KLM, although no practical experience of applying the method.

8.5.2 Procedure

Extended KLMs were produced according to the guidelines as set out in the previous chapter. Static task times were predicted according to the traditional KLM guidelines (Card et al., 1983), with the inclusion of the reach far (Rf) operator developed specifically for the driving context (Green, 2003).

POI Search task

A model was considered for the example task of obtaining route guidance for a journey from the current location to Birmingham Airport. The traditional KLM, for the purposes of predicting static task time, was produced as follows:

Static Task Time = Rf (move hand from lap to control of device – 0.45 seconds) + **M** (1.35 seconds) + **K** (Select ‘Navigate to’ – 0.2 seconds) + **M** (1.35 seconds) + **H** (0.4 seconds) + **K** (Select ‘POI’ – 0.2 seconds) + **M** (1.35 seconds) + **H** (0.4 seconds) + **K** (Select category type – 0.2 seconds) + **M** (1.35 seconds) + **H** (0.4 seconds) + **K** (Select relevant destination – 0.2 seconds) = **7.85 seconds**

This static task time sequence was then used to develop an extended KLM, in order to model task performance under the interrupted vision sequence as experienced by participants in occlusion trials. The model is illustrated in Figure 8.1, where the darker areas represent 1.5-second shutter closed periods, and the lighter areas correspond to 1.5-second shutter open periods. The resulting prediction for TSOT was 7.5 seconds; therefore, R is predicted to be 0.84.

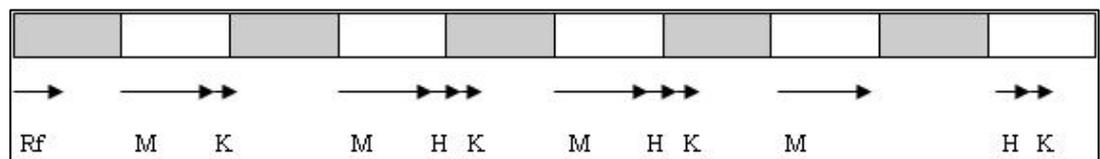


Figure 8.1 Extended KLM for POI search task (shaded areas = 1.5 seconds occlusion; clear = 1.5 seconds vision)

The modeller provided reasoning for some of his decisions made in producing the extended KLM as follows:

The first two H operators are unlikely to require full vision, as they will be automated movements associated with the top-level menus of the device. However, the final H (locating the specific POI name

from a long list – 11 items) will require vision and therefore is modelled to start at the beginning of a shutter open period.

Destination entry task

In this task, the example given was to obtain route guidance to the address '12 Cedar Avenue, Runcorn'. In trials, participants were required to enter the first four letters of the town/city name (Runcorn), then select the full name from a list, select Cedar Avenue from a list, and then enter the number 12. The traditional KLM for this task was as follows:

Static task time = Rf (move hand to system – 0.45 s) + M (1.35 s) + K (select 'Navigate to' – 0.2 s) + M (1.35 s) + H (0.4 s) + K (select 'Address' – 0.2 s) + M (1.35) + 4H (1.6 s) + 4K (enter RUNC – 0.8 s) + M (1.35 s) + H (0.4 s) + K (select 'Runcorn' – 0.2 s) + M (1.35 s) + H (0.4 s) + K (select 'Cedar Avenue' – 0.2 s) + M (1.35 s) + H (0.4 s) + K (enter number – 0.2 s) = **13.6 seconds**

Figure 8.2 illustrates the resulting extended KLM, which predicts a TSOT value of 10.5 seconds; consequently, R is predicted to be 0.77.

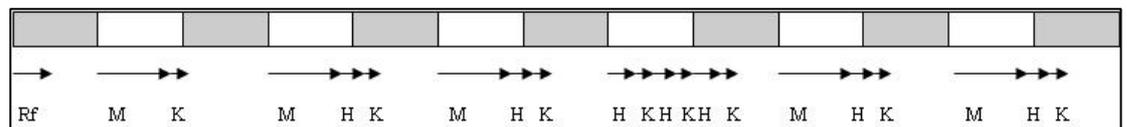


Figure 8.2 Extended KLM for destination entry task (shaded areas = 1.5 seconds occlusion; clear = 1.5 seconds vision)

As with the POI search task, reasoning was provided for decisions made in producing the extended KLM:

It is assumed that only one M is required prior to entering the letters of the city name (as the city is one chunk in working memory). Also, the selection of a specific city and specific street are from short lists (max. 4 items). Therefore, it is assumed that the associated H operator does not require full vision.

Scrolling text tasks

In the simple and complex scrolling text tasks, it was necessary to select variants for modelling that bore closest similarity to the overall mean observed results for the task, due to clear differences observed between variants of a task depending on when the search term first appears on the screen. Thus, models were produced for finding the company code YHR for the short task, and RLQ for the long task, and reading out the associated share prices.

KLM models for the tasks were developed as follows:

- Short task – **static task time** = system response (time for search term to appear on-screen – 10 s) + M (search for item on-screen – 1.35 s) + M (read accompanying share price – 1.35 s) = **12.7 seconds**
- Long task – **static task time** = system response (time for search term to appear on-screen – 5.3 s) + M (search for item on-screen – 1.35 s) + M (read accompanying share price – 1.35 s) = **8.0 seconds**

Figures 8.3 and 8.4 illustrate the resulting extended KLMs for these tasks, predicting a TSOT of 7.35 seconds for the short task and 5.85 seconds for the long task. Consequently, R is predicted to be 0.58 for the short task and 0.73 for the long task.

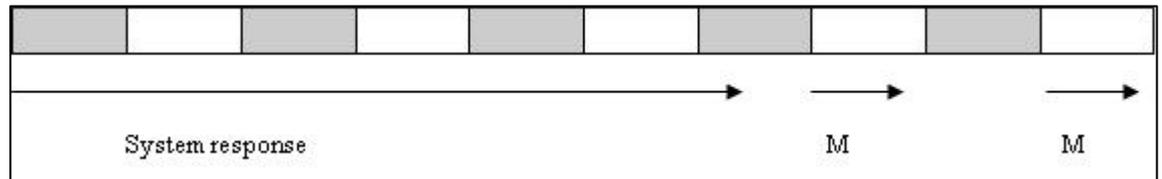


Figure 8.3 Extended KLM for short scrolling text task (shaded areas = 1.5 seconds occlusion; clear = 1.5 seconds vision)

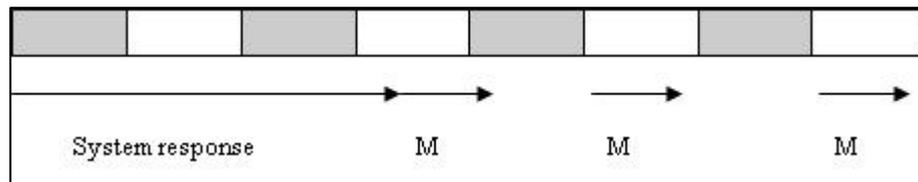


Figure 8.4 Extended KLM for long scrolling text task (shaded areas = 1.5 seconds occlusion; clear = 1.5 seconds vision)

The modeller noted that these tasks were particularly difficult to model, requiring significant assumptions about actual performance to be made. For example, it is necessary to suppose that the target code will be identified as soon as it appears on screen whilst a participant has vision. Furthermore, the mental operator is employed here as a general operator, considering activities such as visual search and speech feedback.

8.6 Findings

Table 8.2 shows a summary of the predicted and observed results. The error rates are low for the first two tasks, but less so for the scrolling text tasks. The modeller concluded that the KLM indicated that differences between the tasks could be attributed to the greater number of stages involved in the destination

entry task, rather than any problem with visual demand, hence a difference between TSOT values and not R. It was also suggested that the lower than expected R-values for the scrolling text tasks could be explained by the significant system response times.

Table 8.2 Summary of observed and predicted results (error rates calculated as difference between 100 and the predicted value as a percentage of observed results, consistent with John et al., (2004a))

Task	Observed TSOT (s)	Predicted TSOT (s)	Error (%)	Observed R	Predicted R	Error (%)
POI Search	7.63	7.5	- 1.70	0.88	0.84	- 4.55
Destination entry	11.77	10.5	- 10.8	0.81	0.77	- 4.94
Scrolling text (short)	10.94	7.35	- 32.7	0.78	0.58	- 25.6
Scrolling text (long)	11.62	5.85	- 49.7	0.87	0.73	- 16.1

8.7 Discussion

8.7.1 Reliability and validity

The results of this evaluation are a first step towards establishing the reliability of the extended KLM, however it should be noted that this is not a full reliability check, since it does not consider factors such as inter-rater reliability, and so on. In essence, the study is a further validation of the extended KLM, with the task models produced independently from the author of this thesis, unlike previous studies. For two of the tasks, results consolidate the findings of the previous two studies (Chapter 6 and Chapter 7), suggesting that the method is valid with these types of interactions. Given that the method was applied independently of the previous studies, it can be inferred that the method is somewhat reliable.

The experiences of the modeller give cause for optimism. Predictions were made independently from user trials, producing results with good error rates, for the destination entry and POI search tasks. For these tasks, the modeller

was able to apply the method without any apparent difficulty, and his decisions regarding occluded task performance appear to conform to the logical assumptions without any significant flaws. However, whilst the method followed by the modeller in this study generally appears sound, the models might be improved with some further analysis. For instance, in the traditional KLM for the destination entry task, the modeller includes only a single keystroke for entering the house number (i.e. only one digit is entered). In the example task, the house number is 12, which suggests there is a need to include further operators for handling this extra input (H for moving to the button, K for pressing). Doing so translates to a static task time prediction of 14.2 seconds, a TSOT value of 11.1 seconds and, consequently, a predicted R-value of 0.78. This leads to a reduction in the prediction error for TSOT to - 5.69% and for R to - 3.70%. It is worth highlighting the fact that this error in the predictions is attributable to an error in the process of creating traditional KLMs, as opposed to the assumptions to be followed as part of the extended KLM.

In the extended KLMs, the modeller made the decision to place the initial operator for both tasks (moving the hand to the device) in an occluded period at the start of the task. Normally, and according to the terms of the ISO standard, user-trial based occlusion studies begin with a shutter open period as soon as task instruction is complete, therefore suggesting that this first operator would be better placed at the beginning of a shutter open period. However, in practical experience, the modeller's assumption in this study may be more in keeping with real world behaviour. In trials, it is quite common for

participants to begin moving their hands towards the system before the investigator has completed task instructions. Given the accuracy of the predictions in this study, there is sufficient cause to believe the models are representative of real world performance.

In contrast to the first two tasks, the scrolling text tasks proved more problematic. Error rates for these tasks surpassed 40% for TSOT and 20% for R. There are two main possibilities for explaining these poor predictions: there are general flaws in the traditional models for these tasks (i.e. static task time error rates are equally poor), or, there are flaws within the extended KLM predictions. The error rate between predicted and observed static task time is 8.73% for the simple task and 41.2% for the complex variation. Therefore, it can be suggested that flaws in the traditional model for the longer task have affected the TSOT prediction, but are less likely to have contributed to the high error rate for the short task. These results highlight an area for further research with the extended KLM; namely, investigations into how specific activities are affected by the interrupted vision conditions of the occlusion paradigm. For example, it may be that new operators are required to consider complex visual searching activities.

The modeller did not experience any significant difficulty with the extended KLM procedure. He did, however, feel that the second logical assumption could be better worded. In its current format, the modeller considered it a little vague. The assumption states:

- An operator that begins in a period of vision can continue into a two second occluded period providing it is not specifically associated with vision, for example, reading information from a display.

The concern is that the assumption is confusing, and could be better expressed. Consequently, the new assumption reads:

- An operator that begins in a period of vision can continue into an occluded period providing it is not specifically associated with vision (e.g. moving towards a known target, completing a key press).

8.7.2 Economy

The purpose of this case study was to examine the reliability of the extended KLM technique; however, it also provides information establishing the method's economy. The modelling process took the modeller little more than half a day. In contrast, the user-trials involved 60 participants, and took place over several weeks. Assuming each trial took about an hour, and that participants were paid for their time at a rate of £10 an hour (a reasonable amount for user-trials), the cost of paying participants alone stands at £600. This is about the equivalent of a day's consultancy fee for the modeller. When staff and set-up costs are added, the user trial procedure will likely be significantly more expensive. The number of participants used here is substantially greater than the ten required by the draft ISO standard, so the costs here are inflated. However, this brief cost evaluation does not take into account the placement of the methods within design cycles. A fundamental benefit of methods that do not require user-trials is that they can be applied in

much earlier design stages, when meaningful design changes are economically feasible (Green, 2003, Salvucci et al., 2005).

8.8 Summary

This chapter presented a case study investigating the reliability of the extended KLM. As noted, there was no consideration of fine-grade reliability characteristics (e.g. inter-rater reliability). Predictions of the TSOT metric made by an independent modeller according to the extended KLMs assumptions and subsequently compared with the results of a large-scale evaluation of the occlusion technique. Error rates were particularly strong for two tasks, further consolidating the validity of the technique in specific circumstances, and providing a first step towards establishing its reliability. However, error rates were higher for scrolling text tasks involving complex visual searching of a display. This discrepancy was not fully explained by flaws in traditional keystroke models, suggesting that new operators or assumptions are required for the extended KLM when tasks require complex visual searching.

In addition, the study provides anecdotal evidence supporting the economic benefits of the technique. Predicted values were produced in under a day, as opposed to the several weeks required to collect empirical data. Finally, through discussion with the modeller, the wording of the second logical assumption was reconsidered to eliminate potential confusion.

9 General Discussion

This chapter presents a general discussion of the key findings that emerge from this research. Within previous chapters, analysis and discussion particular to individual studies was given. As such, it is not the purpose of this chapter to discuss specific results or data at a detailed level. Instead, this chapter will focus on synthesising results, leading to the development of conclusions, to be presented in Chapter 10. Nevertheless, there will be some review and summary of the work that has gone before as part of this discussion, as and when required.

Particular attention is paid to the extended KLM technique, focussing especially on its validity, reliability and economy. In addition, opportunities for further work are identified and discussed. Finally, the discussion focuses on the contributions made to research. This includes: the implications for HCI methods in general, and for the KLM in particular; the use of the extended KLM within other contexts, where the visual demand of an interface may be a concern; and, any design guidance for IVIS tasks that emerges from the experience of conducting this research, and from the theories surrounding the extended KLM.

9.1 *IVIS and driver distraction*

The potential for novel IVIS tasks to dangerously distract drivers is a significant safety concern. As demonstrated in Chapter 2, driver distraction is an established causal factor in road accidents. Significantly, it can be difficult

to legislate against driver distraction, in contrast to other impairments, such as when driving under the influence of alcohol. Distraction, unlike alcohol or fatigue, is likely to affect drivers only intermittently over the course of a journey (Stevens and Minton, 2001). Furthermore, drivers may even choose to engage in distracting tasks during periods where overall accident risk is low, for example when on quiet, straight roads, or when stationary, such as when waiting at traffic lights (Cnossen et al., 2004, Young et al., 2003). As driving processes become more automated, IVIS may even be important in avoiding potential driver under-load (Stevens and Minton, 2001). Consequently, banning drivers from engaging in all IVIS tasks whilst a vehicle is in motion is neither realistic nor practical.

A future research aim for driver distraction in general might be to combine accident analysis studies with evidence from driving performance evaluations, in order to establish the absolute risk posed by IVIS tasks. Risk is a factor of the demand of performing a task, the prevalence of the system amongst the driver population, the frequency of use, and the driving environment at the instant of interaction. Consequently, determining the overall risk of a distraction source must consider exposure to the source, in addition to its distracting effects (Neale et al., 2005). Klauer et al. (2006) go some way to addressing this issue in an analysis of data from the 100-car naturalistic driving study. Their findings suggest secondary task engagement can result in a crash risk up to three times greater than in normal driving depending on task complexity.

For IVIS development, the challenge is to consider potential distraction throughout IVIS design, in order that systems do not pose an unnecessary risk to drivers. Achieving this requires a variety of driver-centred evaluation methods that can be implemented across all stages of the design cycle. As highlighted in Chapter 3, few low-cost methods can be implemented in early design stages. The central aim of the research was to address this issue, and develop a method for evaluating IVIS distraction that is valid, reliable and economical.

9.2 Occlusion

The occlusion technique assesses the visual demand of IVIS tasks, and has been the subject of considerable research in recent years. The technique acts as a surrogate for the running of road or simulator-based user trials, and, in this respect, is considered a cost-effective alternative. High-fidelity driving simulators are expensive to install, maintain and run, contrasting significantly with the relatively cheap equipment necessary to run occlusion trials (Reed and Green, 1999, Young et al., 2003).

Chapter 4 outlined a study assessing the validity of the occlusion technique, with results that were in line with previous work (see Chapter 3). Significantly, the results support the assertion that the TSOT metric is more discerning of the visual demand of a task than measures of static task time (Young et al., 2003, Stevens et al., 2004). The results failed to validate the second occlusion metric, R, as an indicator of tasks that may potentially encourage long single glance durations. It was concluded that R is a

complicated metric, influenced by multiple factors. These include whether or not participants need to repeat task stages because of occlusion periods, whether or not participants need to visually re-orient to a task following an occlusion period, and, whether or not, and to what extent, the task can progress through occlusion periods due to opportunities for blind control. However, the R metric was not considered to be without merit, and was found to relate to some on-road measures such as glance frequency, and the proportion of time spent with eyes-off-road. Recent studies have since suggested there may be a correlation between R and measures of lane keeping performance, and, object and event detection (Bengler, 2006, Tijerina, 2006a).

A concern remains that the occlusion technique relies on measures of performance made through user trials. Such trials typically require well-developed, robust, prototype systems, which are generally available only in latter design stages. Consequently, it is unlikely that results can be used to inform re-design (Green, 2003, Salvucci, 2005). Thus, the economy of the technique is questionable.

9.3 Keystroke Level Model

Performance measures that are normally captured through user trial assessments may alternatively be predicted using modelling techniques. In this respect, the GOMS family of modelling techniques, and in particular, the KLM, have been the focus of significant research since their development in the early 1980s (Card et al., 1983, John, 2003). The KLM, which involves decomposition of a task into its most basic physical and cognitive processes

(known as operators), has been used to predict task time values on a variety of devices, including in-car devices (Manes et al., 1998, Raskin, 2000, Stanton and Young, 2001, John et al., 2004a, John et al., 2004b, Myung, 2004, Luo and John, 2005). Chapter 5 further assessed the validity of the technique in the driving context, with positive results. In summary, results suggest good relative validity of task time predictions, with the use of standard operator values, though care should be taken when input mechanisms differ significantly from those associated with computer systems. Absolute validity is improved through the development of situation-specific operators.

Traditionally, the KLM predicts performance times in single-task contexts. Somewhat similarly, occlusion trials are based on single-task performance under interrupted vision conditions. In this respect, the technique differs from other driver distraction evaluation methods, e.g. the lane-change test, where the focus is on the dual-task of interacting with an IVIS whilst driving. Dual-task methods pose a difficulty, in that a modelling technique has to deal with how and when attention between tasks is interleaved. The theory behind task switching is poorly understood, and is an area for further research (John et al., 2004b). A limitation of the Distract-R system, reviewed in Chapter 3, is that it requires modellers to make decisions about task switching points, a subjective decision (Salvucci et al., 2005). Consequently, performance under the interrupted vision conditions of occlusion may be modelled in a more straightforward fashion than other assessment methods.

9.4 Extended KLM

In extending the KLM to model human performance in occlusion trials, the key challenge was to consider the effects of the interrupted vision sequence on task procedures, and translate these into guidelines for the development of extended keystroke-level models. With the traditional technique, a comparable set of heuristics was provided for the placement of mental operators within a model. The guidelines for the extended KLM, termed logical assumptions, were first developed for the study outlined in Chapter 6 and reinforced in later studies, outlined in Chapter 7 and Chapter 8. In essence, assessments of the validity and reliability of the extended KLM technique are focussed on the corresponding validity and reliability of these guidelines.

9.4.1 Validity

Two studies were conducted to assess the validity of the extended KLM technique (Chapters 6 and 7). The approach, in both cases, was to compare measures generated by extended KLM predictions with observed results. In general, predicted values ranked tasks in a similar fashion to the observed results, with error rates within the 20% to be expected from KLM assessments (John et al., 2004b). Significantly, in cases where error rates were higher, it was concluded that factors other than the extended KLM process had contributed to the increased error. Typically, in these cases, static task time predictions were equally weak, suggesting the discrepancy was between the traditional model and the task as actually performed in the user trials.

Opportunities for improving the validity of the new technique exist, and centre on the development of new, IVIS specific operators. A striking benefit of the

KLM is that it is adaptable to new situations, and new operators can easily be incorporated (John, 2003, Luo and John, 2005). The operators and associated time values used in the studies outlined in this thesis were generally drawn from the original suite developed by Card et al. (1983). These were typically associated with command-line desktop computing tasks, significantly distinct from the touch-screen/GUI based tasks typical of IVIS (John et al., 2004a). Furthermore, whilst all participants were trained to acceptable performance levels for trials their performance is not necessarily comparable with the skilled typists involved in the original assessments (Card et al., 1983). Therefore, there is significant scope for the development of context-specific IVIS operators that consider these factors. Manes et al. (1998) and Green (2003) provide a number of operators specific to in-car systems. Further, Fitts' law can be used to make estimations for actions that are device and context specific, e.g. hand movements (Myung, 2004, Luo and John, 2005).

The most problematic operator is that associated with cognitive processes (M). Further to the issues encountered in the placement of mental operators within a task model, it has been claimed that the operator is too general, since it accounts for several activities, including visual search, recall, decision-making, verification and other perceptual and motor operations (John et al., 2004b). Thus, there is scope for the operator to be adjusted or redefined for the extended KLM. Specifically, it may be worth considering separating the operator to cater for visual and cognitive elements. This is an approach considered by John et al. (2004b), where the 1.35 second mental operator is divided into two operators: a cognitive processing operator, 1200 ms in

duration, and a visual operator of 150 ms duration. This separation is considered sufficient for single task situations (John et al., 2004b).

A potential limitation of the extended KLM is that it is unlikely to predict higher R values. New operators may be required to cater for IVIS tasks where R is likely to exceed one. Such tasks are rarely reported (notable exceptions are Noy et al., 2004 and Angell et al., 2006). There are likely two conditions under which such a result may occur. In the first, system pacing (e.g. system times-out after a given period of no interaction) may mean that participants would miss information due to the visual occlusion procedure. This situation should be catered for by the extended KLM in its current description. The second condition occurs when participants are required to read complex information from a display (e.g. a complicated map, or detailed text), a process that may be disrupted by the visual occlusion procedure. Currently, it is unlikely the extended KLM would accurately model this behaviour, as the process of reading information would simply be separated across vision intervals under the logical assumptions. The scrolling text tasks assessed in Chapter 8 showed this to be a reasonable conclusion, although these tasks did not produce R-values greater than one in user trials. Empirical work is necessary to determine the effect of the occlusion procedure on operators concerned with reading information from displays, which can be translated into new, extended-KLM-specific operators. A starting point would be to evaluate whether the 150 ms visual operator is sufficient to satisfy this limitation (John et al., 2004b).

9.4.2 Reliability

A significant body of research suggests that the KLM, in its original inception, can reliably be repeated to produce valid task time assessments (e.g. Manes et al., 1998, Raskin, 2000, Stanton and Young, 2003, John et al., 2004a, John et al., 2004b, Luo and John, 2005). Therefore, assessments of the reliability of the extended KLM should focus on the three logical assumptions concerning task performance under the interrupted vision sequence of occlusion trials. The wording of these assumptions was carefully considered, in order to reduce the likelihood for multiple interpretations. However, the decisions made by a modeller guided by these assumptions are ultimately subjective. A reliable KLM prediction is dependent on a rigorous and consistent application of the heuristics for the placement of mental operators (John et al., 2004b), and the same might be said of the extended KLM and its associated guidelines.

An independently conducted extended KLM study was presented in Chapter 8, demonstrating that the technique appears to be reliable. A HCI specialist, with theoretical knowledge of the extended KLM, produced predictions for the TSOT and R metrics associated with four IVIS tasks. These predictions were subsequently compared with the results of an occlusion study, conducted separately by a group of independent researchers, involving sixty participants. Absolute error rates in this study were low for two of the tasks, and, once a discrepancy in the traditional KLM for one task was rectified, these fell below 10%. These results are particularly encouraging given the independence of the modelling and user-trial processes, and the comparatively large participant sample. The remaining tasks' higher error rates suggested a limitation of the extended KLM. Specifically, the technique currently struggles to model tasks

that require participants to read complex visual information (see above). The experiences of the modeller were also used to redefine one of the logical assumptions, in order to avoid confusion.

However, one study alone does not fully establish reliability, and there is scope for significant further work in this area. A difficulty in conducting reliability studies with the extended KLM lies in the complex background to the technique. Novice modellers must be educated in several concepts before they may confidently apply the method, including the traditional KLM, the theory behind the occlusion technique, and, the extended KLM's logical assumptions. If participants in a reliability study misunderstand aspects of the first two concepts, this may translate to inaccuracies in their subsequent TSOT predictions. Therefore, reliability studies must be very carefully designed, to ensure that the focus remains on the reliability of the logical assumptions. Encouragingly, evidence suggests that the traditional KLM is well understood and easily learnt by novices (Stanton and Young, 2001). Consequently, a successful, more comprehensive reliability study might reduce concerns over the technique's portability.

Perhaps somewhat unfairly, the KLM has a reputation as being difficult to apply (Dix et al., 2003, John, 2003, John et al., 2004b). This explains, to a certain extent, why the technique remains marginalised in comparison with user-trial methods (John, 2003). In order to address this concern, John et al. (2004a, 2004b) describe the development of CogTool, a software application that automatically generates keystroke-level task time predictions from

demonstrations of tasks on HTML storyboards. There are two fundamental benefits of the tool: it makes the method more accessible, and it removes a degree of subjectivity from the KLM procedure. This second point is crucial for further establishing the reliability of the technique. If the subjective aspects of the technique are removed, reliability is assured. Consequently, the development of a similar tool for the extended KLM is desirable.

9.4.3 Economy

The economic value of an evaluation method is dependent on two components, and is not simply the cost of implementation. Measures of economy must also consider at which design stages a technique can be implemented. Thus, in general, the earlier the method can be applied in the design process, the more economical it is, as design changes can realistically be implemented. By predicting occlusion metrics through keystroke modelling, the intention was to maximise the benefits of the occlusion technique whilst reducing the costs. Amongst the perceived costs of the technique are the limitations imposed by performance-based user trials. Such trials require robust, reliable prototype systems, which are typically only available in latter design stages; by which point design changes are less practical (Green, 2003, Salvucci, 2005). Consequently, the value of the metrics associated with the technique is diminished.

Whilst the economy of the extended KLM was not specifically assessed through the course of this work, it can be inferred that the new technique significantly improves on the cost of running user-based occlusion trials.

Evaluating the costs involved in the study outlined in Chapter 8 demonstrated, to some extent, that this is a fair assumption. The cost of developing the models and predictions was equivalent to the money required for participant time alone in the user-trials, let alone researcher's wages, equipment costs, participant recruitment, and so on. However, the true economic benefits of the technique can only be demonstrated when applied in a real design cycle. In contrast to user trials, an extended KLM assessment can be undertaken with only a design concept and outline of task procedures. Consequently, designs can be altered in accordance with the predictions, with little difficulty. This process would be particularly straightforward if predictions are automated through software such as CogTool (John et al., 2004b), since designers would quickly be able to evaluate a variety of task procedures, and interface design configurations, and select the most favourable for further development.

9.5 Summary of extended KLM future work

The above discussion focussing on the validity, reliability and economy of the extended KLM draws out some of the future research challenges that emerge from this work. These issues, along with others, may be summarised as follows.

- *Scope* – Currently, the extended KLM uses simple assumptions regarding task performance in occlusion conditions, in combination with traditional KLM operators, and appears to be valid for particular tasks, e.g. discrete menu-based interactions. However, more visually complex tasks may be less effectively considered, thus further work may wish to focus on quantifying

elements such as the time taken to visually reorient to a display following interruptions.

- *Enrichment* – The task level detail provided by an extended KLM assessment provides an opportunity to develop new metrics associated with the visual demand of engaging with an interface. Specifically, extended KLM allows analysts to quantify precisely the elements of a task that afford blind interaction with a device. Thus, future research may wish to consider what additional metrics, beyond TSOT and R, may be generated from extended KLM analyses, and the potential value of these for IVIS designers.
- *Reliability* – The traditional KLM has been repeatedly demonstrated as a reliable technique when rigorously applied (e.g. Manes et al., 1998, Stanton and Young, 2003). Given this, reliability assessments of the extended KLM should therefore focus on the reliability of the logical assumptions and their application by independent modellers. Further, future work may wish to evaluate the technique in a realistic design cycle, where it is used to assess low-fidelity prototypes.
- *Implementation* – It has been noted that the traditional KLM, despite its enthusiastic supporters, has failed to be adopted by practitioners to the same degree as user trial and heuristic methods. This is in part attributed to the effort that needs to be invested in learning how to apply the technique. A potential way to address this concern lies in the development of software

tools that remove the practicalities of the technique from the users' control, and the CogTool application does this for the traditional KLM. Thus, a future research aim for the extended KLM would be to extend this open-source application, or develop an original application, in order that practitioners can mock-up interface designs, demonstrate task procedures and automatically generate extended KLM analyses.

9.6 Contribution to research

The GOMS and KLM theories were first presented to the research community over 25 years ago, and have since been the subject of substantial research (see John, 2003). In contrast to user-trial methods, there is a general feeling that the techniques, whilst valid and reliable, are difficult to learn and apply, although the KLM is considered the simplest to apply GOMS technique, and the easiest to explain and justify to developers (John, 2003). However, it is fair to say that the techniques, including the KLM, are little used in the practical development of computer interfaces, despite significant research interest (John, 2003). This is an undesirable situation, since the proven benefits of the techniques are substantial (Dix et al., 2003, John, 2003).

This research presented a novel development of the KLM, extending the technique to model performance under interrupted vision conditions. Whilst several studies have focussed on the development of new KLM operators, the author is unaware of any studies that use KLM techniques to model performance in conditions other than the full-vision task-user environment of the original method. As such, the thesis makes an important contribution to

KLM and GOMS research, in further demonstrating the flexibility of the technique. In an age when novel computing devices are becoming increasingly prolific and ubiquitous, it is important to establish that techniques, such as the KLM, which was initially developed for command-line desktop computing interfaces, remain valid and relevant.

9.6.1 The extended KLM in the context of other evaluation methods

Chapter 1 introduced a framework into which evaluation methods might be placed according to the concepts of validity, reliability and economy. Existing evaluation methods were placed within this framework in Chapter 3, with the suggestion that no method satisfied all conditions. This was attributed to the reliance on user-trials with a focus on human performance measures; consequently, few methods could be described as economical. The extended KLM, however, might tentatively be placed within the centre of the framework as shown in Figure 9.1, since it appears to be valid, reliable, and, in contrast to user-trial methods, economical.

In some respects, the framework is a little simplistic. For example, it does not give an indication of scale for each criterion; the PDT, on-road evaluation and simulator study are all valid, but how valid? To address this concern, in Chapter 3, methods were plotted by validity and economy. Figure 9.2 incorporates the extended KLM within this alternative framework. In this case, it is made explicit that the extended KLM (or occlusion) does not provide results at the same level of validity as, for example, an on-road evaluation,

though this does not mean they are not valid results. However, it demonstrates that the technique is significantly more economical than such studies.

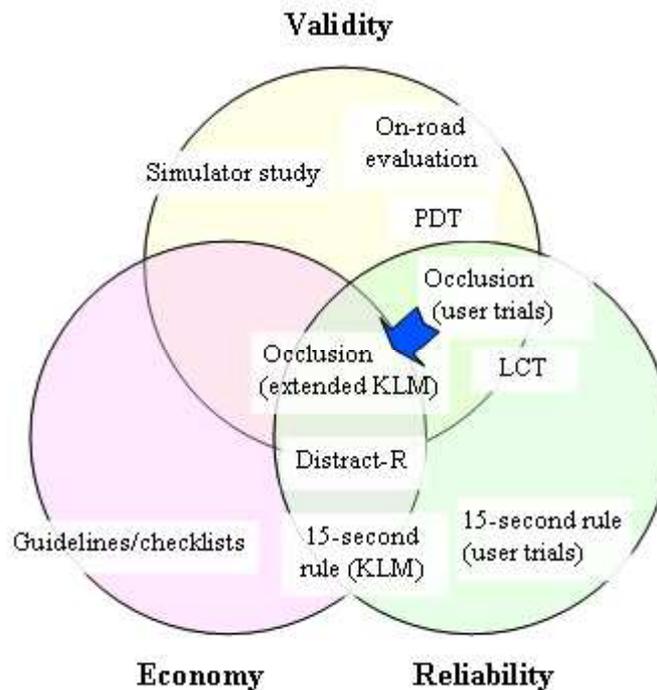


Figure 9.1 Extended KLM within the evaluation framework

As with Figure 9.1, Figure 9.2 is a simplified representation also. The figure does not consider the differing scopes of the evaluation methods. Standardised methods (e.g. occlusion, 15-second rule, LCT) are particularly focussed; two such studies are likely to cost about the same, and the results will generally be valid and comparable. In contrast, other methods, such as simulator studies, are vague; simulators can range from low-cost, basic systems utilising desktop computers and games-wheels, to high fidelity, realistic installations, incorporating real-vehicle cabs and controls, immersive virtual reality environments, and motion bases (Young et al., 2003). This concept of differing method scope is not represented in the diagram.

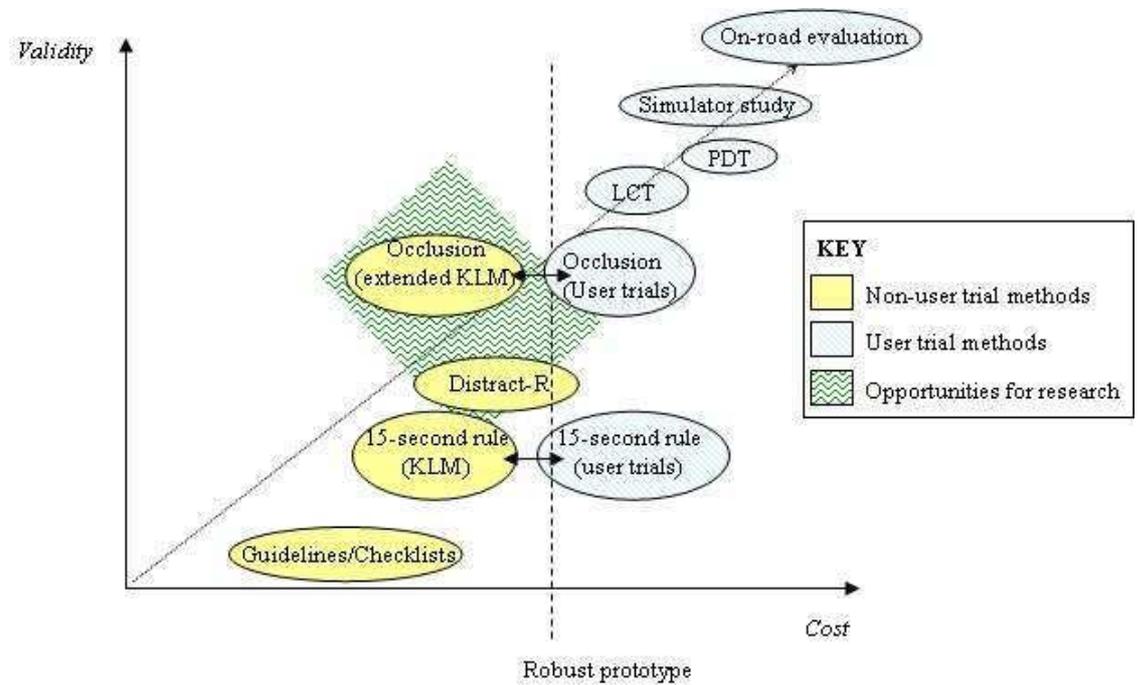


Figure 9.2 Evaluation methods according to validity and economy

An alternative categorisation of evaluation methods is presented in Angell et al. (2006). Here the authors developed a toolkit of evaluation methods, placing modelling techniques in early development stages, occlusion in early physical prototypes, and simulator evaluations in more advanced prototype stages (see Figure 9.3). As has been stated, the fundamental benefit of the extended KLM technique is that it allows evaluation of visual demand, as normally assessed through occlusion trials, to take place in much earlier design stages, thus maximising the value of the associated metrics. Therefore, the extended KLM could be placed within the pre-prototype phase in Angell et al.'s toolkit.

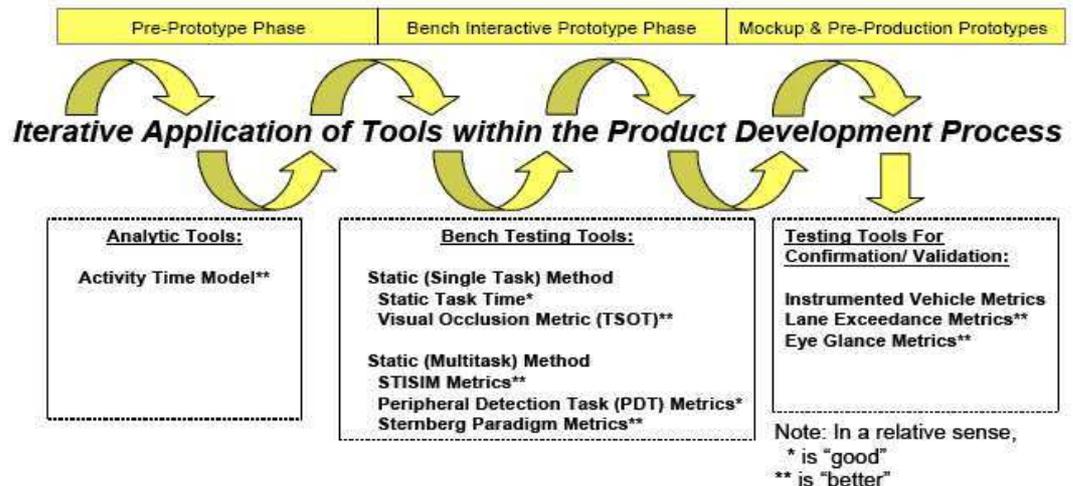


Figure 9.3 Driver Workload Metrics Toolkit (Angell et al., 2006)

In considering the range of available evaluation methods in a global sense, a further research opportunity is identified; efforts might focus on how methods are, and should be, used in the real development of IVIS. Research in the field has matured to the extent that methods are significantly well established. The occlusion technique, for example, is standardised (ISO, 2007), and others, such as the LCT, are in the process. At first glance, it might appear that methods are competing for dominance. However, it is likely that each provides a unique view of distraction potential. In this respect, the design and development of an IVIS could utilise a variety of evaluation methods throughout its lifetime, in order to consider its potential for driver distraction from multiple perspectives. Potentially, a design-cycle evaluation program might be as follows:

- *Start/Early design* – adherence to guidelines and checklists in order to constrain the design based on established notions of safety

- *Low-fidelity prototype phase* – use of extended KLM to distinguish between different design configurations based on likely visual demand
- *Medium-fidelity (bench) prototype phase* – use of lab-based techniques (e.g. PDT, LCT) to assess system(s)
- *High-fidelity prototype phase* – simulator and road studies

The extended KLM can be applied within the early stages of design, requiring that the practitioners have information regarding its dimensions, the characteristics of the controls, its likely placement in the vehicle, and the task configurations. By allowing for occlusion metrics to be developed at this stage (low-fidelity prototypes), other methods may be applied in place of traditional occlusion trials, thus enriching the overall evaluation process over the design lifecycle.

However, the above evaluation framework is merely one suggestion. A variety of techniques are available for evaluating distraction and research establishing how and when these should be applied is timely. The workload metrics toolkit is unique in considering a range of methods and their placement within a design cycle (Angell et al., 2006). Future research might consult with design and development professionals to validate, and/or redefine, this toolkit.

9.6.2 Use of the extended KLM in other contexts

The focus of this research, and in the development of the extended KLM, was on driver distraction and IVIS task design. Whilst this is possibly the most relevant context for the method, and was the motivating factor for the research,

it may also be applicable in other environments. Sophisticated computing technology is increasingly unconfined to desktop situations; for example, the advertising for the Nokia N95 mobile phone states boldly ‘It’s what computers have become’.¹⁴ This situation presents new challenges for interface and system designers. In particular, there may be a growing number of systems where interface designers wish to reduce the level of visual demand that may be placed on users. This may be due to the physical constraints of the device, such as restricted space for displays and controls, or, the context within which the system will be used, such as wearable computing devices, where users may be engaged in simultaneous, real-world tasks (Baber, 2001). In these instances, the extended KLM (and the occlusion technique) may be useful in ensuring interfaces are designed to suit their purpose.

9.6.3 Implications for IVIS design

The purpose of this research was not specifically to consider the design of IVIS tasks; however, lessons can be drawn from the presented findings and discussions. The extended KLM procedure is based on theories concerning the way tasks are performed under occlusion conditions, resulting in an assessment of potential visual demand. An extended KLM assessment demonstrates, in particular, opportunities where users may be able to progress with a task without visual attention. In general, these opportunities concern either cognitive actions (i.e. preparing for the next task stage) or simple, repetitive actions. A proven benefit of the original KLM is the qualitative information it provides about task behaviour (John, 2003). A close examination of a KLM

¹⁴ Nokia N95, <http://nokia.co.uk/n95>

can highlight recurring sub-procedures that might be combined or shortened, for instance (John, 2003). Consequently, the extended KLM could be used to identify opportunities to structure tasks for enhanced interrupted performance.

This research supports the notion that IVIS designers should focus on developing short tasks. In the majority of cases, tasks with long TSOT values have comparably more stages than those with acceptable values. In cases where long tasks are unavoidable, a more favourable TSOT assessment is the result of significant opportunity for blind control. In this case, design should focus on the development of interfaces that support this activity. Such interfaces might utilise audio feedback mechanisms and haptic controls. Another approach may seek to assist users in managing a long task in a series of manageable interactions. This will likely involve some degree of trade-off with overall task length; however, the increased duration may be attributable to extra cognitive operators, which can be readily completed without visual attention. In the study outlined in Chapter 7, the destination entry task for System A had a greater static task time and TSOT value than System B. This was principally due to the extra task steps required by the system, and system response delays. However, System A's R-value was lower than System B's since system responses regularly fell into occlusion periods, and participants were also able to mentally prepare during these periods.

9.7 Summary

This chapter has presented a general discussion of the main findings emerging from the empirical and literature research of the previous chapters. Analysis of

results and discussion of specific study details was presented within each chapter; therefore, this chapter has focussed on synthesising results, providing the basis for general conclusions. Discussion focussed on the validity, reliability and economy of the extended KLM in particular, since the development of this method has been the focus of the research. Subsequently, discussion considered the implications for, and contribution made to, research in a wider sense. This included the relevance of the research for KLM, GOMS and HCI methods, the role of the extended KLM within a framework of driver distraction IVIS assessment methods, the application of the method in other computing task contexts, and, the implications of the technique for IVIS task design. The main conclusions are summarised in Chapter 10.

10 Conclusions

The aim of this thesis was to develop a method for evaluating the potential visual distraction, or visual demand, of IVIS tasks that was valid, economical and reliable. Three further objectives were identified:

1. To investigate the issue of driver distraction, clarify the definition of the term and gauge its importance as a factor in accident causation
2. To assess through studies and available literature the range of methods currently available for measuring distraction potential, focusing on their validity, reliability and economy
3. To develop a new economical method, or expand/adapt existing method(s), for measuring distraction potential and test its validity and reliability

This chapter presents the main conclusions in relation to these three objectives that emerge from this research.

10.1 Distraction and accidents (Objective 1)

In addressing the first objective, the relationship between road accidents and driver distraction was reviewed. It was demonstrated that driver distraction plays a significant role in accident causation, and, therefore, potentially distracting in-car tasks are a safety concern. However, accident research can be problematic; studies have used a variety of data sources, and may be difficult to compare. There is also the concern that many distraction-related accidents are not reported in traditional accident databases, as demonstrated by

the significantly greater proportion of accidents attributed to distraction within the 100-car naturalistic driving study (Neale et al., 2005, Klauer et al., 2006). Whilst studies generally suggest distraction is a contributory factor for between 2% and 25% of accidents, the real figure may be considerably larger.

Another fundamental problem is that the term driver distraction is somewhat vague, and may in some cases be confused with the broader category of inattention. In this respect, a comprehensive definition was proposed, stating that driver distraction is:

- *Delay by the driver in the recognition of information necessary to safely maintain the lateral and longitudinal control of the vehicle (the driving task)*
- *Due to some event, activity, object or person, within or outside the vehicle,*
- *That compels or tends to induce the driver's shifting attention away from fundamental driving tasks,*
- *By compromising the driver's auditory, biomechanical, cognitive or visual faculties, or combinations thereof.*

Understanding the nature of driver distraction, and considering the associated safety concerns, provided the grounding for the more substantial work of the thesis.

10.2 Evaluation methods (Objective 2)

Various evaluation methods are available for assessing new in-vehicle systems in terms of their potential for driver distraction. Methods include standardised procedures, such as the occlusion technique and the 15-second rule, and less well defined techniques, utilising driving simulators or instrumented vehicles. The majority of available methods are assessed through user trials with a focus

on measures of human performance. A fundamental issue here is that such trials can only take place in more advanced design stages, by which time it may not be economically viable to make adjustments to the interface design (Green, 2003, Salvucci, 2005).

In this respect, the experimental work focussed on the development of a non-user trial method. Non-user techniques include qualitative assessments and performance modelling approaches. Performance modelling allows for quantitative predictions of task times to be made in early development stages. The KLM has previously been used to model performance in IVIS tasks, specifically for 15-second rule assessments. Therefore, it was concluded that a suitably adapted modelling approach might be feasible with other user-trial methods.

10.3 Development of extended KLM (Objective 3)

The occlusion technique was selected as the basis for a new, task modelling technique. Unlike other evaluation methods, the technique does not require assessment of interfaces in dual-task conditions, and was therefore considered suitable for modelling. The technique was evaluated as part of an on-road evaluation. The results supported the findings of other research, which suggest that the technique, and its associated metrics, provides a valid indication of the visual demand of IVIS tasks (see Chapter 3). In particular, the technique is more sensitive than another single-task evaluation technique, the 15-second rule, which is based on static task times. In line with other research, the TSOT metric was found to be of most interest in evaluating visual demand, with a

secondary metric, R, providing extra information about task complexity and visual demand.

The KLM task analysis technique, developed by Card et al. (1983), was extended in order to model performance in the interrupted vision conditions of the occlusion paradigm. As such, the technique could be used to make predictions of the TSOT metric. Since the original technique has long been considered to make valid and reliable predictions of static task time, predictions of TSOT and static task time can be used to calculate a predicted value for the second occlusion metric R.

The production of an extended KLM involves two stages: firstly, tasks are modelled according to the rules of the traditional KLM, in order to provide static task time predictions; secondly, task models are assessed against the interrupted vision conditions of the occlusion paradigm, in order to predict TSOT. Modelling occluded task performance requires the analyst to consider how users will progress in tasks when their vision is occluded. In order to support this process, three logical assumptions were presented:

- **Assumption 1:** During 1.5-second periods of vision the operator sequence can progress without interruption;
- **Assumption 2:** An operator that begins in a period of vision can continue into an occluded period providing it is not specifically associated with vision (e.g. moving towards a known target, completing a key-press);

- **Assumption 3:** An operator can only begin in an occluded period when vision is not required at any point in its duration (e.g. a key-press when the finger is already on the target).

Chapters 6 and 7 sought to establish the validity of the extended KLM technique, with Chapter 8 evaluating reliability and economy. The results of these studies were generally positive, with indications that the extended KLM may be considered valid, reliable and economical. However, there is also considerable scope for further work. In particular, further assessments should be made of the technique's reliability. The critical factor in such studies is ensuring that modellers fully understand the concepts behind the traditional KLM and the occlusion technique, such that studies can focus on the reliability of the logical assumptions. Another approach might be to develop software tools that automate extended KLM production, and the CogTool application, which can be used to generate traditional KLMs, may be an appropriate starting point (John et al., 2004a, John et al., 2004b). Further work may also focus on the development of novel operators specific to the use of IVIS, and may even consider how to redefine the mental operators in order to better represent task performance under occlusion conditions. Specifically, it is a concern that, in its current format, the technique may not predict R-values greater than one. Therefore, there is scope for a new operator that accounts for the time necessary to visually reorient to a task following interruptions when displays are sufficiently complex.

Throughout the course of this research the author has sought to publicise the development of the extended KLM and study details have been presented to the CHI (Computer-Human Interaction) community at their annual conference, the ITS community at their world congress, and ergonomics practitioners at the annual society conference. A key motivation for this dissemination is that the method should be taken up by other researchers, tested and further developed accordingly. The overall reception for the technique has been enthusiastic, with reviewers for CHI (the community most familiar with KLM) noting that it presents a novel development for the KLM, and that there is nothing like it in the existing literature. In particular, Stuart Card developer of the original KLM and co-author of seminal text *The Psychology of Human-Computer Interaction* remarked on the usefulness of the technique, particularly on the concept of taking two established methods from different communities (occlusion – transport safety, KLM – HCI) and combining them for the mutual benefit of both.

10.4 Summary

This research presented the development of an extended KLM technique for use in assessing the visual demand of computing tasks. The research was inspired by the need for low-cost, quantitative methods for assessing the potential distraction of in-vehicle information and communication systems. Current evaluation methods tend to favour performance-focussed user-trial approaches. Consequently, the majority are not considered economically effective. The occlusion technique was considered suitable for performance modelling with a suitably adapted KLM approach. This research has presented

the development and validation of the extended KLM technique, as well as investigating its reliability and economy. The new technique is considered a valuable contribution to research. It further demonstrates the flexibility, adaptability and relevance of KLM approaches. Secondly, it is considered useful beyond the context of in-vehicle technology, particularly with the rise in the popularity of more portable computing devices. Finally, implementation of the technique can provide designers with useful qualitative information about interface configurations in the context of visual demand.

References

- AAM (2003) Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-vehicle Information and Communication Systems. Version 3.0. Washington, D.C., June 2003, Alliance of Automobile Manufacturers.
- ANGELL, L., AUFLICK, J., AUSTRIA, P. A., KOCHHER, D., TIJERINA, L., BIEVER, W., DIPTIMAN, T., HOGSETT, J. & KIGER, S. (2006) Driver Workload Metrics Task 2 Final Report. Report No. DOT HS910 635. Washington, D.C., USA, NHTSA.
- ASOH, T. & IIHOSHI, A. (2006) Occlusion method to evaluate visual distraction to be caused by using car navigation systems. *Driver Metrics Workshop*. Ottawa, Canada. Presentation available at: <http://ppc.uiowa.edu//drivermetricsworkshop>.
- BABER, C. (2001) Wearable computers: a human factors review. *International Journal of Human-Computer Interaction*, 13(2), 123-145.
- BAUMANN, M., KEINATH, A., KREMS, J. F. & BENGLER, K. (2004) Evaluation of in-vehicle HMI using occlusion techniques: experimental results and practical implications. *Applied Ergonomics*, 35(3), 197-205.
- BAUMEISTER, L., JOHN, B. E. & BYRNE, M. (2000) A comparison of tools for building GOMS models. CHI 2000, ACM Conference on Human Factors in Computing Systems. *CHI Letters*, 2(1), 502-509.
- BBC NEWS (2007) Drivers face new phone penalties. BBC News, 27th February 2007 ed.: available at <http://news.bbc.co.uk/1/hi/uk/6399407.stm>, accessed on 7th March 2007. London.
- BEIRNESS, D. J., SIMPSON, H. M. & DESMOND, K. (2002) The Road Safety Monitor 2002: Risky Driving. Ottawa, Canada, Traffic Injury Research Foundation.
- BENGLER, K. (2006) Occlusion: a method to assess visual demand and interruptability. *Driver Metrics Workshop*. Ottawa, Canada. Presentation available at: <http://ppc.uiowa.edu//drivermetricsworkshop>
- BISHOP, R. (2005) *Intelligent Vehicle Technology and Trends*, London, Artech House Books.
- BLAAUW, G. J. (1982) Driving experience and task demands in simulator and instrumented car: a validation study. *Human Factors*, 24, 473-486.
- BURCH, D. (2002) The mobile phone report: a report on the effects of using a hand-held and a hands-free mobile phone on road safety. Croydon, UK, Direct Line Insurance.
- BURNETT, G. E. & PORTER, J. M. (2002) An empirical evaluation of the use of distance versus landmark information within the Human-Machine Interface to vehicle navigation systems. IN DE WAARD, D., BROOKHUIS, K. A., WEIKART, C. M. & TOFETTI, A. (Eds.) *Human Factors in Transportation, Communication and the Workplace*. The Netherlands, Shaker Publishing, 49-64.

- BURNETT, G. E., SUMMERSKILL, S. J. & PORTER, J. M. (2004) On-the-move destination entry for vehicle navigation systems: unsafe by any means? *Behaviour and Information Technology*, 23(4), 265-272.
- BURNS, P. C., HARBLUK, J. L., GO, E. & MORTON, A. H. (2006a) Counting clicks and scrolls: a simple approach to quantifying driver distraction risk. *IEA 2006: 16th World Congress on Ergonomics*. Maastricht, The Netherlands.
- BURNS, P., HARBLUK, J., TRBOVICH, P. & LOCHNER, M. (2006b) Evaluating tasks with the LCT. *Driver Metrics Workshop*. Ottawa, Canada. Presentation available at: <http://ppc.uiowa.edu//drivermetricsworkshop>
- BURNS, P. C. (2006) Transport Canada: R occlusion research: road safety and motor vehicle regulation. Detroit, USA, Presentation to AAM, 19th June 2006.
- CARD, S. K., MORAN, T. P. & NEWELL, A. (1983) *The Psychology of Human-Computer Interaction*, London, Lawrence-Erlbaum.
- CARSTEN, O. M. J., GROEGER, J. A., BLANA, E. & JAMSON, A. H. (1997) Driver performance in the EPSRC Driving Simulator (LADS): A validation study. Final Report for EPSRC project GR/K56162. Leeds, University of Leeds.
- CHIANG, D. P., BROOKS, A. M. & WEIR, D. H. (2001) An experimental study of destination entry with an example automobile navigation system (SAE paper 2001-01-0810). *Intelligent Vehicle Initiative (IVI): Technology and Navigation Systems*. Society of Automotive Engineers. 113-123.
- CLARKE, D. D., WARD, P., BARTLE, C. & TRUMAN, W. (2005) An in-depth study of work-related road traffic accidents. Final report for the Department for Transport. School of Psychology, University of Nottingham, March 2005.
- CNOSSEN, F., MEIJMAN, T. & ROTHENGATTER, T. (2004) Adaptive strategy changes as a function of task demands: a study of car drivers. *Ergonomics*, 47(2), 218-236.
- COOPER, P. J. & ZHENG, Y. (2002) Turning gap acceptance decision making: impact of driver distraction. *Journal of Safety Research*, 33, 321-335.
- CURRY, R., GREENBERG, J. & BLANCO, M. (2002) An alternative method to evaluate driver distraction. *Intelligent Transportation Society of America's Twelfth Annual Meeting and Exposition*.
- DFT (2006) Mobile phones and driving: frequently asked questions. Available at: http://www.dft.gov.uk/stellent/groups/dft_rdsafety/documents/page/dft_rdsafety_025216.hcsp.
- DFT (2007) Road Casualties Great Britain: 2005. UK, Department for Transport. <http://www.dft.gov.uk/pgr/statistics/datatablespublications/accidents/causaltiesgbar/roadcasualtiesgreatbritain2005>.
- DINGUS, T. A. (1987) Attentional demand evaluation for an automobile moving-map navigation system. Blacksburg, VA, USA, Virginia Polytechnic Institute and State University.

- DINGUS, T. A., ANTIN, J. F., HULSE, M. C. & WIERWILLE, W. W. (1989) Attentional demand requirements of an automobile moving-map navigation system. *Transportation Research A*, 23(4), 301-315.
- DIX, A., FINLAY, J., ABOARD, G. D. & BEALE, R. (2003) *Human Computer Interaction: 3rd Edition*, USA, Pearson-Prentice Hall.
- EBY, D. W. & KOSTYNIUK, L. P. (1999) An on-the-road comparison of in-vehicle navigation assistance systems. *Human Factors*, 41(2), 295-311.
- ENDSLEY, M. R. (1988) Design and evaluation for situation awareness enhancement. *Human Factors Society 32nd Annual Meeting*. Santa Monica, CA, USA, 97-101.
- ENDSLEY, M. R. & GARLAND, D. J. (2000) *Situation Awareness Analysis and Measurement*, UK, Lawrence Erlbaum Associates.
- ENGSTRÖM, J. & MARKKULA, G. (2006) Effects of visual and cognitive load on the lane change test - preliminary results. *Driver Metrics Workshop*. Ottawa, Canada, Presentation available at <http://ppc.uiowa.edu/drivermetricsworkshop/>.
- ESBJÖRNSSON, M., JUHLIN, O. & WEILENMANN, A. (2007) Drivers using mobile phones in traffic: an ethnographic study of interactional adaptation. *Int. J. Human-Computer Interaction*, 22, 1-2, 37-58.
- FARBER, E., FOLEY, J. & SCOTT, S. (2000) Visual attention design limits for ITS in-vehicle systems: the Society of Automotive Engineers standard for limiting visual distraction whilst driving. *Transportation Research Board Annual Meeting*. Washington, DC, USA.
- FOLEY, J. & HUDAK, M. (1996) Route guidance system evaluation - Part 2: post-drive results. Technical report for the 'Product of the Future Fund', Ford Motor Company, USA.
- GALER, M. (1994) Driver vehicle interaction and human factors. IN PAUWELUSSEN, J. P. & PACEJKA, H. B. (Eds.) *Smart Vehicles*. UK, Swets and Zeitlinger.
- GODLEY, S. T., TRIGGS, T. J. & FILDES, B. N. (2001) Driving simulator validation for speed research. *Accident Analysis and Prevention*, 34, 589-600.
- GOODMAN, M. J., BENTS, F. D., TIJERINA, L., WIERWILLE, W. W., LERNER, N. & BENEL, D. (1997) An investigation into the safety implications of wireless communication in vehicles. Washington, DC, USA, Department of Transportation, NHTSA.
- GOUJON, S. (2001) First evaluation of possible occlusion technique criteria. *Workshop on Occlusion*. Torino, Italy.
- GREEN, P. (1999a) The 15-second rule for driver information systems. *Intelligent Transportation Society of America - 9th Annual Meeting and Exposition ("New Thinking in Transportation")*. Washington D.C., USA, ITS America.
- GREEN, P. (1999b) Estimating compliance with the 15-second rule for driver-interface usability and safety. *Human Factors and Ergonomics Society 43rd Annual Meeting*. Santa Monica, California.
- GREEN, P. (1999c) Visual task demands of driver information systems. Technical Report UMTRI-98-16. Ann Arbor, Michigan, USA, University of Michigan Transportation Research Institute.

- GREEN, P. (2000) Crashes induced by driver information systems and what can be done to reduce them. SAE Paper 2000-01-C008. Washington, DC, Society of Automotive Engineers.
- GREEN, P. (2003) Motor vehicle driver interfaces. IN JACKO, J. A. & SEARS, A. (Eds.) *The Human Computer Interaction Handbook*. London, Lawrence-Erlbaum Associates.
- GREEN, P., HOEKSTRA, E. & WILLIAMS, M. (1993) Further on-the-road tests of driver interfaces: examination of a route guidance system and car phone. Report No. UMTRI-93-35. Ann Arbor, USA, University of Michigan Transportation Research Institute.
- HAIGNEY, D. E., TAYLOR, R. G. & WESTERMAN, S. J. (2000) Concurrent mobile (cellular) phone use and driving performance: task demand characteristics and compensatory processes. *Transportation Research Part F*, 3, 113-121.
- HAIGNEY, D. E. & WESTERMAN, S. J. (2001) Mobile phone use and driving: a critical review of research methodology. *Ergonomics*, 44, 132-143.
- HARBLUK, J. L., NOY, Y. I. & EIZENMAN, M. (2002) The impact of cognitive distraction on driver visual behaviour and vehicle control. TP No. 13889 E. Canada, Transport Canada.
- HARMS, L. & PATTEN, C. (2003) Peripheral detection as a measure of driver distraction. A study of memory-based versus system-based navigation in a built-up area. *Transportation Research Part F*, 6, 23-36.
- HASHIMOTO, K. & ATSUMI, B. (2001) Study of occlusion technique for making the static evaluation method of visual distraction. *Occlusion Workshop*. Turin, Italy.
- HORBERRY, T., STEVENS, A., ROGERS, A., COTTER, S. & BURNETT, G. E. (2007a) WP2: Consideration of different IVIS tasks: expert evaluation, keystroke level task analysis and social acceptability ratings. Interim Report. Crowthorne, Berkshire, UK, TRL Limited.
- HORBERRY, T., STEVENS, A., ROGERS, A., COTTER, S., ROBBINS, R. & BURNETT, G. E. (2007b) WP1: Effects of Age: Occlusion further research. Interim Report. Crowthorne, Berkshire, UK, TRL Limited.
- ISO (2004a) Road vehicles - Ergonomic aspects of transport information and control systems - Occlusion method to assess visual distraction due to the use of in-vehicle information and communication systems. ISO TC 22/SC13 N 763 R. International Standards Organisation.
- ISO (2004b) Road vehicles - Ergonomic aspects of transport information and control systems - Occlusion method to assess visual distraction due to the use of in-vehicle systems. ISO/CD 16673. International Standards Organisation
- ISO (2005) Road vehicles - Ergonomic aspects of transport information and control systems - Occlusion method to assess visual distraction due to the use of in-vehicle information and communication systems. ISO DIS 16673. International Standards Organisation.
- ISO (2006) Road vehicles - Ergonomic aspects of transport information and control systems - Occlusion method to assess visual distraction due to use of in-vehicle information and control systems. ISO DIS 16673.2. International Standards Organisation.

- ISO (2007) Road vehicles - Ergonomic aspects of transport information and control systems - Occlusion method to assess visual demand due to the use of in-vehicle systems. ISO International Standard 16673.
- JANSSEN, W. (2000) Driver distraction in the European statement of principles on in-vehicle HMI: a comment. Internet forum on driver distraction. <http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/Welcome.htm>
- JOHN, B. E. (1995) Why GOMS? *Interactions*. October 1995 ed.
- JOHN, B. E. (2003) Information processing and skilled behaviour. IN CARROLL, J. M. (Ed.) *HCI Models, Theories and Frameworks: Toward a Multidisciplinary Science*. London, Morgan-Kaufmann.
- JOHN, B. E., PREVAS, K., SALVUCCI, D. D. & KOEDINGER, K. (2004a) Predictive human performance modelling made easy. *CHI 2004*. ACM Press, New York.
- JOHN, B. E., SALVUCCI, D. D., CENTGRAF, P. & PREVAS, K. (2004b) Integrating models and tools in the context of driving and in-vehicle devices. *International Conference on Cognitive Modelling 2004*. Pittsburgh, PA, USA.
- KARBASSIOUN, D. (2004) A critical review of the Keystroke Level Model (Card, Moran & Newell, 1983) as a method for evaluating user interfaces in a driving context. MSc thesis, UK, University of Nottingham.
- KARLSSON, R. & FICHTENBERG, N. (2001) How different occlusion intervals affect total shutter open time. *Workshop on Occlusion*. Torino, Italy.
- KARPUTER (2006) <http://www.karputer.co.uk>.
- KLAUER, C. (2006) Naturalistic driving data: one of many tools in a transportation researcher's tool box. *Driver Metrics Workshop*. Ottawa, Canada, <http://ppc.uiowa.edu/drivermetricsworkshop/>.
- KLAUER, S. G., DINGUS, T. A., NEALE, V. L., SUDWEEKS, J. D. & RAMSEY, D. J. (2006) The impact of driver inattention on near-crash/crash risk: an analysis using the 100-car naturalistic driving study data. Report No. DOT HS 810 594. Washington DC, NHTSA.
- KOEDINGER, K. R., ALEVEN, V. & HEFFERNAN, N. (2003) Toward a rapid development environment for cognitive tutors. *Proceedings of AI-ED 2003*. IOS Press.
- LAI, F. C. H. (2005) Driver attentional demand to dual task performance. Leeds, University of Leeds.
- LAM, T. (2002) Distractions and the risk of car crash injury: the effect of drivers' age. *Journal of Safety Research*, 33, 411-419.
- LANDAY, J. A. & MYERS, B. A. (1995) Interactive sketching for the early stages of user interface design. *CHI 95*. ACM Press, New York.
- LANSDOWN, T. C., BURNS, P. C. & PARKES, A. M. (2004) Perspectives on occlusion and requirements for validation. *Applied Ergonomics*, 35, 3, 225-232.
- LEE, J. D., CAVEN, B., HAAKE, S. & BROWN, T. L. (2001) Speech-based interaction with in-vehicle computers: the effects of speech-based E-mail on drivers' attention to the roadway. *Human Factors*, 45, 631-639.

- LESCH, M. F. & HANCOCK, P. A. (2004) Driving performance during concurrent cell-phone use: are drivers aware of their performance decrements? *Accident Analysis and Prevention*, 36, 471-480.
- LUO, L. & JOHN, B. E. (2005) Predicting task execution time on handheld devices using the Keystroke-Level Model *CHI 2005*. Portland, Oregon, ACM, New York.
- MANES, D., GREEN, P. & HUNTER, D. (1998) Prediction of destination entry and retrieval times using keystroke level models. Michigan, USA, UMTRI.
- MATTES, S. (2003) The lane-change-task as a tool for driver distraction. *Proceedings of IGfA 2003*.
- MATTES, S. (2006) The lane-change test. *Driver Metrics Workshop*. Ottawa, Canada. Presentation available at: <http://ppc.uiowa.edu/drivermetricsworkshop>.
- MCKNIGHT, A. J. & MCKNIGHT, A. S. (1993) The effect of cellular phone use upon driver attention. *Accident Analysis and Prevention*, 25, 259-265.
- MICHON, J. A. (1993) *Generic Intelligent Driver Support: A Comprehensive Report on the GIDS Project*, London, Taylor and Francis.
- MILGRAM, P. (1987) A spectacle-mounted Liquid-Crystal Tachistoscope. *Behaviour Research Methods, Instruments and Computers*, 19(5), 449-456.
- MILGRAM, P. (2005) Translucent Technologies, <http://home.ca.inter.net/~milgram/>
- MOLLENHAUER, M. A., HULSE, M. C., DINGUS, T. A., JAHNS, S. K. & CARNEY, C. (1997) Design decision aids and human factors guidelines for ATIS displays. IN NOY, Y. I. (Ed.) *Ergonomics and Safety of Intelligent Driver Interfaces*. New Jersey, USA, Lawrence-Erlbaum Associates.
- MYUNG, R. (2004) Keystroke-level analysis of Korean text entry methods on mobile phones. *International Journal of Human-Computer Studies*, 60, 545-563.
- NATIONAL STATISTICS (2006) Adult mobile phone ownership or use: by age, 2001 and 2003: social trends 34, available at <http://www.statistics.gov.uk/STATBASE/ssdataset.asp?vlnk=7202>, accessed on 10th August 2006.
- NEALE, V. L., DINGUS, T. A., KLAUER, S. G., SUDWEEKS, J. & GOODMAN, M. J. (2005) An overview of the 100-car naturalistic study and findings. Paper no. 05-0400. Washington DC, USA, National Highway Traffic Safety Administration.
- NIIYA, K. (2000) Trial evaluation results on navigation operation using occlusion technique. *ISO/TC22/SC13/WG8 Visual Distraction Task Force*. Toyota Motor Corporation.
- NOY, Y. I. (1997) *Ergonomics and Safety of Intelligent Driver Interfaces*, London, Lawrence Erlbaum Associates.
- OLSSON, S. & BURNS, P. C. (2000) Measuring distraction with a peripheral detection task. Online paper available at: <http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/PDF/6.pdf>. Washington DC, USA, NHTSA.

- PARKES, A. (1997) The role of standards for in-vehicle MMI. IN NOY, Y. I. (Ed.) *Ergonomics and Safety of Intelligent Driver Interfaces*. New Jersey, USA, Lawrence Erlbaum Associates.
- PARKES, A. (2005) Technology and distracted driving: technology-push, demand-pull and the consequences of poor design. *International Conference on Distracted Driving*. Toronto, Canada.
- PETTITT, M. A., BURNETT, G. E., BAYER, S. & STEVENS, A. (2006a) Assessment of the occlusion technique as a means for evaluating the distraction potential of driver support systems. *IEE Proceedings on Intelligent Transport Systems*, 4, 259-266.
- PETTITT, M. A., BURNETT, G. E. & KARBASSIOUN, D. (2006b) Applying the keystroke level model in a driving context. IN BUST, P. D. (Ed.) *Contemporary Ergonomics: Proceedings of the Ergonomics Society Annual Conference 2006*. Cambridge, UK, Taylor and Francis.
- PETTITT, M. A., BURNETT, G. E. & STEVENS, A. (2005) Defining driver distraction. *Proceedings of 12th ITS World Congress*. San Francisco, USA, ITS America.
- PETTITT, M. A., BURNETT, G. E. & STEVENS, A. (2006c) Extending the keystroke level model (KLM) to assess the visual demand of an in-vehicle information system (IVIS). *Proceedings of 13th World Congress on Intelligent Transport Systems 2006*. London, UK, Ertico.
- PETTITT, M. A., BURNETT, G. E. & STEVENS, A. (2007) An extended keystroke level model (KLM) for predicting the visual demand of in-vehicle information systems. *CHI'2007*. San Jose, California, USA, ACM.
- POLLARD, E. (1994) *The Oxford Paperback Dictionary*, Oxford, UK, Oxford University Press.
- PREECE, J., ROGERS, Y. & SHARPE, H. (2002) *Interaction Design*, New York, J. Wiley & Sons.
- RANNEY, T., MAZZAE, E., GARROTT, R. & GOODMAN, M. J. (2000) NHTSA Driver distraction research: past, present and future. NHTSA Internet forum on driver distraction. Available at: <http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/PDF/233.PDF>.
- RASKIN, J. (2000) *The Humane Interface*, London, Addison-Wesley.
- REDELMEIER, D. A. & TIBSHIRANI, R. J. (1997) Association between cellular telephone calls and motor-vehicle collisions. *The New England Journal of Medicine*, 336(7), 453-458.
- REED, M. P. & GREEN, P. A. (1999) Comparison of driving performance on-road and in a low-cost simulator using a concurrent telephone dialling task. *Ergonomics*, 42(8), 1015-1037.
- ROCKWELL, T. H. (1988) Spare visual capacity in driving revisited: new empirical results for an old idea. IN GALE, A. G., FREEMAN, M. H., HASLEGRAVE, C. M., SMITH, P. & TAYLOR, S. P. (Eds.) *Vision in Vehicles II*. UK, Elsevier Science.
- ROSPA (2002) The risk of using a mobile phone while driving. Online paper, available at: <http://www.rospa.org.uk/pdfs/road/mobiles/report.pdf>. Royal Society for the Prevention of Accidents.
- SAE (2000) SAE J2364 Recommended practice: navigation and route guidance function accessibility while driving. Society of Automotive Engineers.

- SALVUCCI, D. D. (2001) Predicting the effects of in-car interface use on driver performance: an integrated model approach. *International Journal of Human-Computer Studies*, 55, 85-107.
- SALVUCCI, D. D. (2006) Modelling driver behaviour in a cognitive architecture. *Human Factors*, 48, 362-380.
- SALVUCCI, D. D., BOER, E. R. & LIU, A. (2001) Toward an integrated model of driver behaviour in a cognitive architecture. *Transportation Research Record*, 1779, 9-16.
- SALVUCCI, D. D., CHAVEZ, A. K. & LEE, F. J. (2004) Modelling effects of age in complex tasks: a case study in driving. *Proceedings of the 26th Annual Conference of the Cognitive Science Society*. Chicago, USA, August 5-7, 2004.
- SALVUCCI, D. D., ZUBER, M., BEREGOVAIA, E. & MARKLEY, D. (2005) *Distract-R*: Rapid prototyping and evaluation of in-vehicle interfaces. *CHI'2005*. Portland, USA, April 2005, ACM Press.
- SANDERS, M. S. & MCCORMICK, E. J. (1992) *Human Factors in Engineering and Design: 7th Edition*, New York, McGraw-Hill, Inc.
- SENDERS, J. W., KRISTOFFERSON, A. B., LEVISION, W., DIETRICH, C. W. & WARD, J. L. (1966) An investigation of driver information processing Washington, DC, US Department of Commerce.
- SHUTKO, J. & TIJERINA, L. (2006) Eye glance behaviour and lane exceedences during driver distraction. *Driver Metrics Workshop*. Ottawa, Canada. Presentation available at: <http://ppc.uiowa.edu/drivermetricsworkshop>
- SIMPSON, H. M. (2005) Distracted driving: how can we prove it's a problem? . *Proceedings of International Conference on Distracted Driving*. Toronto, Canada, October 2-5, 2005.
- STANTON, N. A. & YOUNG, M. S. (2001) Giving ergonomics away? The application of ergonomic methods by novices. *Applied Ergonomics*, 34, 479-490.
- STEVENS, A., BYGRAVE, S., BROOK-CARTER, N. & LUKE, T. (2004) Occlusion as a technique for measuring in-vehicle information system (IVIS) visual distraction: a research literature review. Crowthorne, Berkshire, UK, Transport Research Laboratory (TRL).
- STEVENS, A. & MINTON, R. (2001) In-vehicle distraction and fatal accidents in England and Wales. *Accident Analysis and Prevention*, 33, 539-545.
- STEVENS, A., QUIMBY, A., BOARD, A. & RAI, G. (2002) Development of human factors guidelines for in-vehicle information systems. *ITS World Congress*. Chicago, USA, ITS America.
- STRAYER, D. (2005) Driven to distraction. *International Conference on Distracted Driving*. Toronto, Canada, October 2-5, 2005.
- STUTTS, J. C., FEAGANES, J., RODGMAN, E., HAMLETT, C., MEADOWS, T., REINFURT, D., GISH, K., MERCADANTE, M. & STAPLIN, L. (2003) Distractions in everyday driving. Washington, DC, AAA Foundation for Traffic Safety.
- STUTTS, J. C., REINFURT, D. W., STAPLIN, L. & RODGMAN, E. A. (2001) The role of driver distraction in traffic crashes. Washington, D.C., USA, AAA Foundation for Traffic Safety.

- TASCA, L. (2005) Driver distraction: towards a working definition. *International Conference on Distracted Driving*. Toronto, Canada, October 2-5, 2005.
- THIMBLEBY, H. (1984) Generative user-engineering principles for user interface design. IN: Proceedings of INTERACT '84, 1st IFIP International Conference on Human-Computer Interaction, 661-666
- THIMBLEBY, H., CAIRNS, P. & JONES, M. (2001) Usability analysis with Markov models. *ACM Transactions on Computer-Human Interaction*, 8, 2, 99-132.
- TIJERINA, L. (2006a) CAMP driver workload metrics project: correlation of the R metric with driving performance and prior prediction. *Driver Metrics Workshop*. Ottawa, Canada. Presentation available at: <http://ppc.uiowa.edu/drivermetricsworkshop>
- TIJERINA, L. (2006b) A measurement systems analysis of total shutter open time (TSOT). *Driver Metrics Workshop*. Ottawa, Canada. Available at: <http://ppc.uiowa.edu/drivermetricsworkshop/>.
- TIJERINA, L., JOHNSTON, S., PARMER, E., WINTERBOTTOM, M. D. & GOODMAN, M. J. (2000) Driver distraction with wireless communications and route guidance systems. Report no. DOT HS 809-069. Washington, DC, USA, National Highway Traffic Safety Administration.
- TIJERINA, L., PARMER, E. & GOODMAN, M. J. (1998) Driver workload assessment of route guidance system destination entry while driving: a test track study. *Proceedings of the 5th ITS World Congress*. Seoul, Korea.
- TIMES (2006a) Some drivers will pay less under 'intelligent' congestion charging. *The Times*. June 27th 2006 ed. London.
- TIMES (2006b) Speed kills, but not as much as bad driving, figures show. *The Times*. 28th September 2006 ed. London.
- TREAT, J. R., TUMBUS, N. S., MCDONALD, S. T., SHINAR, D., HUME, R. D., MAYER, R. E., STANSIFER, R. L. & CASTELLAN, N. J. (1979) Tri-level study of the causes of traffic accidents: final report. Report No. DOT-HS-034-3-535-79-TAC(S). Bloomington, Indiana, USA, Institute for Research in Public Safety.
- TRIGGS, T. J. (1996) Driving simulation for railway crossing research. *7th International Symposium on Railroad-Highway Grade Crossing Research and Safety - Getting Active at Passive Crossings*. Monash University, Australia.
- VAN WINSUM, W., MARTENS, M. & HERLEND, L. (1999) The effect of speech versus tactile driver support messages on workload, driver behaviour and user acceptance. TNO-report TM-99-C043. Soesterberg, Netherlands, TNO.
- VESTE, H. T. & BRUYAS, M. P. (2006) Adaptation of LCT measurement using an adapted curve for each subject. *Driver Metrics Workshop*. Ottawa, Canada, <http://ppc.uiowa.edu/drivermetricsworkshop/>.
- WALKER, G. H., STANTON, N. A. & YOUNG, M. S. (2001) Where is computing driving cars? *International Journal of Human-Computer Interaction*, 13(2), 203-229.
- WANG, J., KNIPLING, R. R. & GOODMAN, M. J. (1996) The role of driver inattention in crashes: new statistics from the 1995 crashworthiness

- data system. *40th Annual Conference for the Advancement of Automotive Medicine*. Vancouver, British Columbia, Canada.
- WEINER, B. (1972) *Theories of motivation: from mechanism to cognition*, Chicago, USA, Rand McNally.
- WEIR, D. H., CHIANG, D. P. & BROOKS, A. M. (2003) A study of the effect of varying visual occlusion and false duration conditions on the driver behaviour and performance of using secondary task human-machine interface. *2003 SAE World Congress*. Detroit, Michigan, USA, ADE Technical Papers Series 2003-01-0128.
- WICKENS, C. D., LEE, J. D., LIU, Y. & GORDEN BECKER, S.E. (2004) *An Introduction to Human Factors Engineering: 2nd Edition*, New Jersey, Pearson Prentice Hall.
- WIERWILLE, W. W. (1993) An initial model of visual sampling of in-car displays and controls. IN GALE, A. G., BROWN, I. D., HASLEGRAVE, C. M., KRUYSSSE, H. W. & TAYLOR, S. P. (Eds.) *Vision in Vehicles IV*. Amsterdam, The Netherlands, Elsevier Science Publishers B.V.
- YOUNG, K., REGAN, M. & HAMMER, M. (2003) Driver distraction: a review of the literature. Report No. 206. Australia, Monash University.
- ZWANEVELD, P. J., VAN AREM, B., BASTIAENSEN, E. G. H. J., SOETEMAN, J. J., FRÉMONT, G., BÉLARBI, F., ULMER, B., BONNET, C. & GÖLLIGER, H. (1999) Deployment scenarios for advanced driver assistance systems. Report Intro/VK 1999-07. Delft, The Netherlands, TNO.

Appendix A – Prevalence of driver distraction: results from the work related road traffic accident database.

This appendix presents a study of the involvement of distraction in work-related road traffic accidents. The intention is to provide new results regarding accident statistics and driver distraction, in order to support our understanding of the relationship between distraction and automobile accidents. This information is useful in providing background to the main research presented in the thesis. However, the detail presented within this review is not considered essential within the main body of the work. Relevant results were discussed in Chapter 2 as necessary.

Work related road traffic accident database

The work-related road traffic accident database covers accidents that occurred between the years 1996 and 2004 inclusive in the Midlands region of the United Kingdom. The database contains information on 2,114 accidents over the period. Each record is generated from police road accident files of accidents involving drivers and others using the roads in connection with their work (Clarke et al., 2004). The information contained within each file varies, and is dependent on the severity of the accident and any subsequent legal proceedings. The minimum amount of information for each accident includes the time, date, location, weather conditions, a brief accident account and other items (Clarke et al., 2004). Files are categorised as being ‘A’ class or ‘B’ class. ‘A’ class files are the most detailed, and contain a range of further items,

such as maps and interviews with witnesses, helping to provide a full account of the complex circumstances of the accident (Clarke et al., 2004).

The data is reasonably specific and no formal attempt will be made to generalise the results to the whole driving population of the United Kingdom. All accidents in the database involve at least one work-related vehicle, although fault or blame does not necessarily lie with the work-related driver. Work-related road traffic accidents are of interest for research into driver distraction for several reasons. Drivers on work-related activities may potentially be exposed to more in-vehicle technology than the overall driving population, with communication devices such as CB and taxi radios prevalent in addition to mobile phones. Furthermore, such drivers may be more motivated to use these devices due to work pressures. With regards to distraction from technology-based sources, work-related accidents may give an indication of future trends as in-vehicle technology becomes increasingly ubiquitous.

In separate research involving the same database of accidents it was concluded that ‘work-related accidents are not fundamentally different in their causal structure to any other road accidents’ (Clarke et al., 2004, p. 40). Furthermore, the errors and violations made by drivers in the database were not significantly different from those made by the general driving population; though work-related drivers have more opportunity to make them (Clarke et al., 2004).

Method

A working definition of driver distraction, as given in Chapter 2, must consider four aspects: the distinction between distraction and inattention; that distraction sources may be both internal and external to the vehicle; the four types of distraction; and, the relationship between a distraction source and the driving task. Analysis of results from the work-related accident database will be structured around these four aspects. This provides further support for the definition presented in Chapter 2, and a coherent structure for the accident analysis.

Results and discussion

Distraction and Inattention

Distraction is listed as a primary or contributory factor in 109 accidents in the work-related road traffic accident database. Of these, two are cases where the distracted party was a pedestrian, and are therefore excluded from analysis. Therefore, 5.1% of accidents in the database list distraction as a primary or contributory factor. Closer inspection of the individual accident reports suggests that up to 29 of these accidents are more likely due to inattention, e.g. daydreaming, fatigue, rather than distraction specifically. The prose accounts of these accidents generally suggest that the drivers were momentarily ‘distracted’ by thoughts of events in their lives other than the driving task at hand. The lack of a definable event or occurrence in these situations leads to the assertion that these accidents are more reasonably categorised as inattention, or ‘looked but did not see’. Efforts should be made within accident reporting to differentiate between distraction and inattention more clearly, in order to assist in post-crash analysis of accident causation. Misleading

classification of accidents can be conducive to inaccurate statistics regarding accident causation, and is potentially harmful for research initiatives and legislative bodies.

Removing the 29 miscategorised cases gives a revised figure of 3.7% of work-related accidents due, in some part, to distraction. This figure is relatively small in comparison with other studies. It is possible that this is due to particular characteristics of the driving population considered here. Professional drivers may be more resistant to the effects of distraction due to better experience, or training, for example. When making comparisons between their analysis of fatal accidents in England and Wales and previous studies, Stevens and Minton (2001) felt this issue might have been a factor. HGV drivers, in particular, are less likely to be exposed to passengers as a distraction source (Stevens and Minton, 2001).

The importance of differentiating between distraction and inattention accidents is made even clearer when accidents are analysed by severity. Within the database, accidents are categorised as 'slight', 'severe' or 'fatal'. Slight accidents involve vehicle damage, which can be substantial, but no serious injury that requires hospitalisation. Often, accidents of this type involve at least one vehicle being written off. Severe accidents usually involve substantial damage to vehicles and some degree of injury to victims involving hospitalisation. Accidents are categorised as fatal when a death occurs because of the accident. The distribution of severity amongst all accidents is shown in

Figure A.1, amongst all distraction related accidents in Figure A.2, and amongst distraction related accidents, eliminating those considered due to inattention, in Figure A.3.

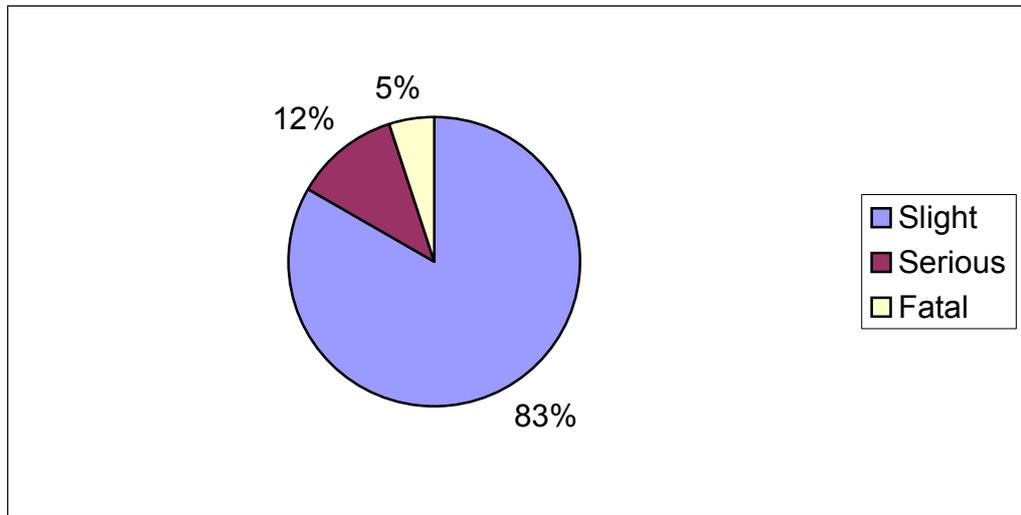


Figure A.1 All accidents by severity (n=2,114)

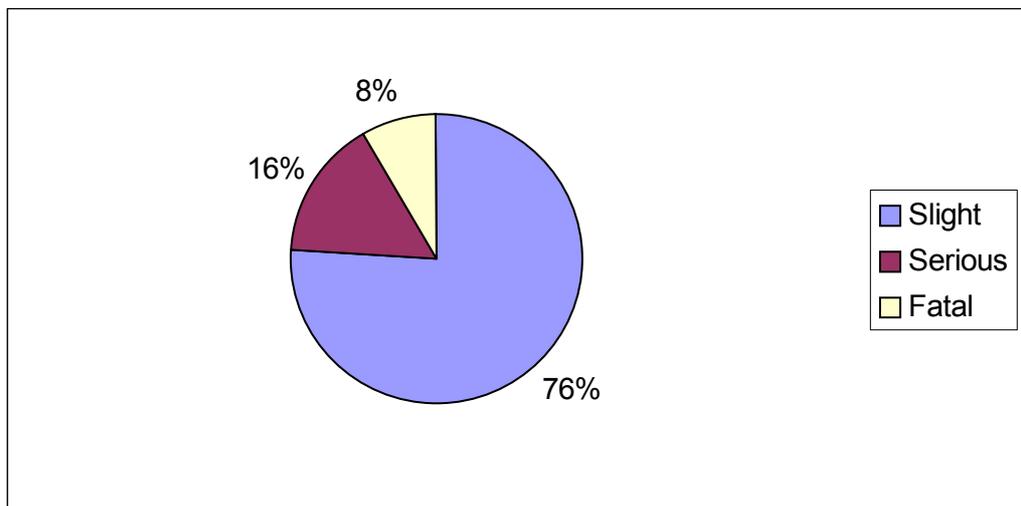


Figure A.2 Distraction related accidents by severity (n=107)

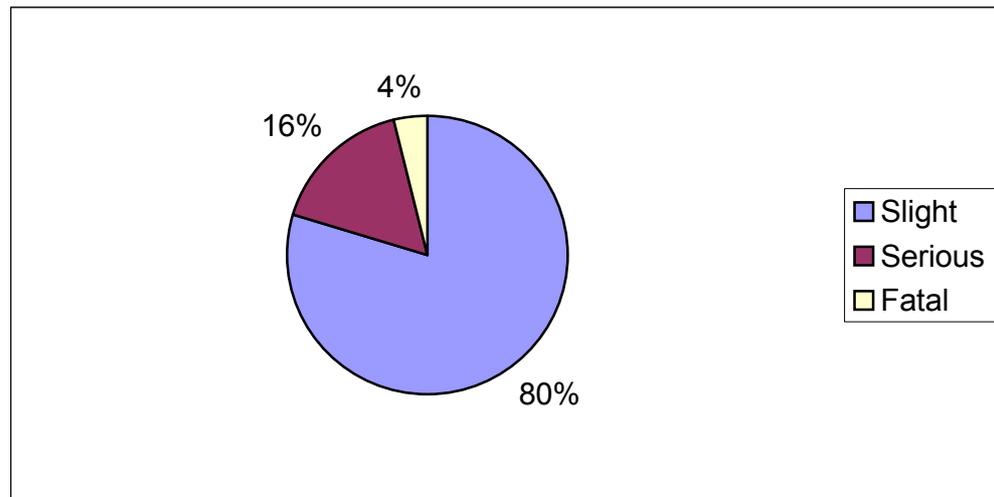


Figure A.3 Distraction related accidents, removing those due to inattention, by severity (n=78)

As can be seen in the figures, removing inattention accidents makes an important difference in the data. Across all accidents, 83% are slight in severity. For those accidents categorised as due, in part, to distraction, 76% are slight. A higher proportion of distraction related accidents appear to be either serious or fatal (24%), when compared with all accidents in the database (17%). Removing those accidents felt to be due to inattention and not distraction reduces these differences. The proportions of slight, serious and fatal accidents are now much closer to those of all accidents (see figures A.1 and A.3). This is further evident when the proportion of distraction related accidents in each category of severity is examined (see figures A.4 and A.5).

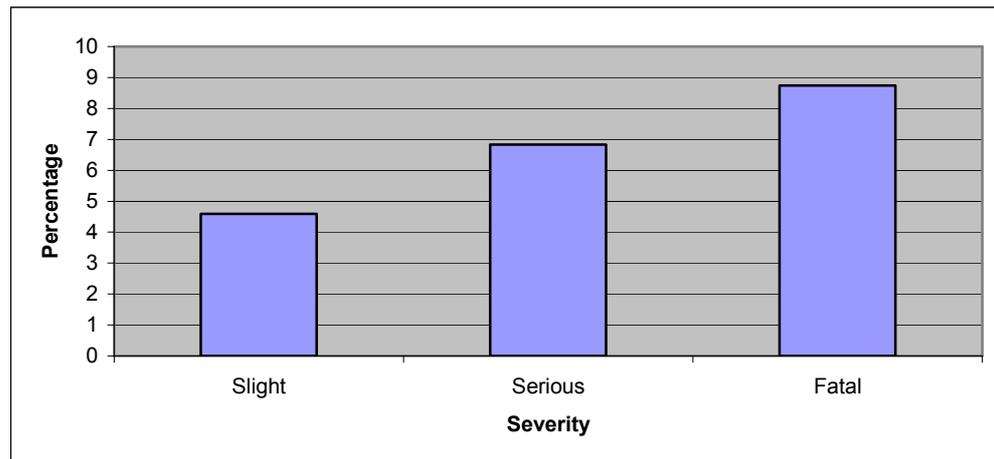


Figure A.4 Percentage of distraction accidents in each severity category (n=107)

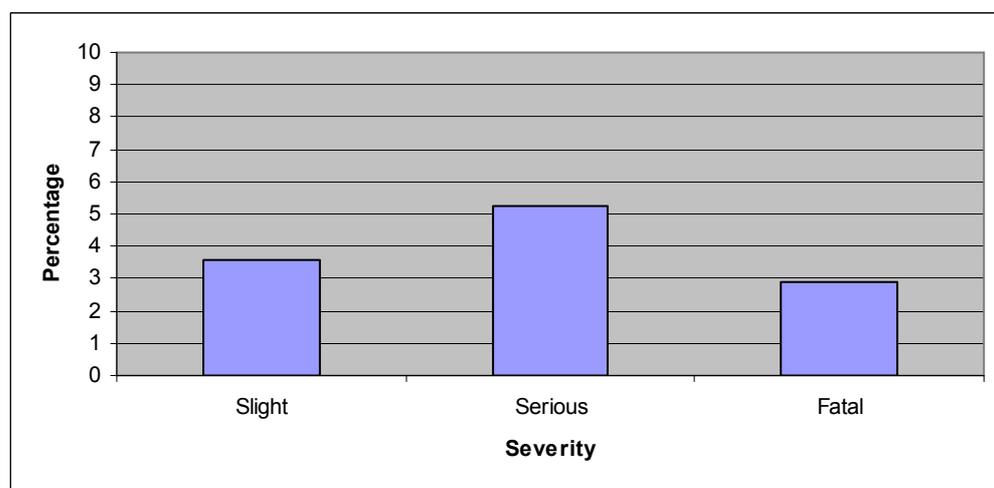


Figure A.5 Percentage of distraction accidents in each severity, removing inattention accidents (n=78)

When all accidents coded with distraction as a contributing factor are considered, distraction is accountable for 4.6% of all slight accidents, 6.8% of serious accidents and 8.8% of fatal accidents (figure A.4). This is of concern, as it suggests that distraction plays a larger role the more severe the accident. However, removing those accidents considered more likely to be due to inattention, the figures change to 3.6% for slight accidents, 5.2% for serious accidents and 2.9% for fatal accidents (figure A.5). The increase in proportion between slight and severe accidents may, in part, be due to the high number of large vehicles, such as HGVs and buses, in the database. Such vehicles have

the potential to cause more damage than if the same accident were to occur with two cars.

Internal and external distraction

Figure A.6 shows the distribution of distraction sources in the accident database. Internal distraction, both from technology sources and non-technology sources, play the greatest role. However, external distraction is also prominent. It has previously been suggested that all forms of external distraction are non-driver initiated (Lam, 2002). However, according to the analysis of the results here, this assertion may be viewed as inaccurate. The prose accounts of accidents given in the database suggest that many accidents where external distraction is a factor are the result of drivers who are lost or unfamiliar with their location, or searching for visual cues such as road signs and street names at the time of the collision. In these situations, it could be suggested that the driver has given priority to their route finding task at the expense of the safe control of their vehicle. This supports other research, such as Cnossen et al. (2004), which concluded that drivers are more likely to accept detriments to their driving performance when a secondary task serves an overall driving task goal. For example, good route finding performance is essential to the overall goal of reaching the destination.

Other instances of external distraction fit more easily with the classification *non-driver initiated*, and include drivers being distracted by other road events and failing to properly attend to the forward view as a result. Examples of other external distractions include cyclists, police vehicles, events off road and

parked vehicles. The distracting effect of searching for road and street signs in unfamiliar environments may be overrepresented in the database as it covers work-related accidents, and drivers on work-related business may drive in unfamiliar locations more regularly than the driving population as a whole.

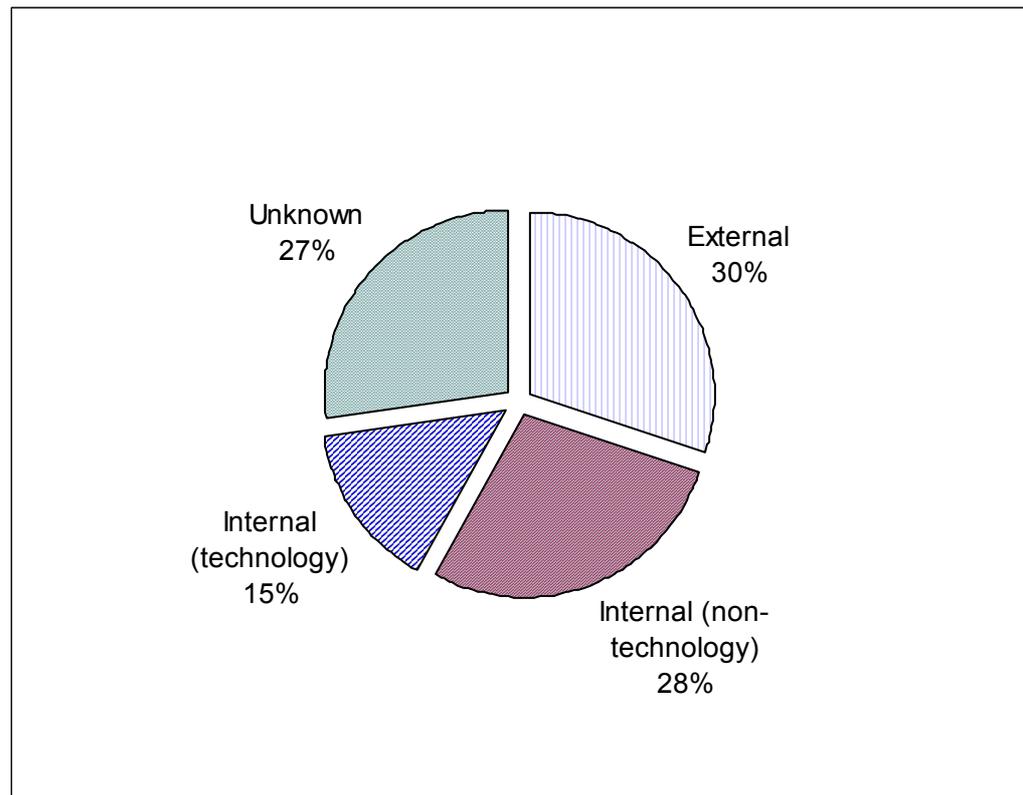


Figure A.6 Distraction sources

The various sources of distraction are summarised in Table A.1. The data has been ranked to show the greatest contributors. Worryingly, mobile phones are the second greatest contributor of internal distraction, involved in seven accidents. One fatal accident in the database involved a driver using a hands-free mobile. The work-related accident database covers accidents between 1996 and 2004, and is therefore more up-to-date than sources used in other post-crash analysis (Wang et al., 1996; Stevens and Minton, 2001). Since the work of previous studies, mobile phone ownership has increased, and

correspondingly, the potential safety risks of using a mobile phone whilst driving may be demonstrated in this data. Similarly, whereas previous research has found CD/Radio/Cassette players to be a large source of distraction (e.g. Stutts et al., 2001), this is not found to be the case here. This may be due to improvements in technology, e.g. ‘finger-tip’ steering wheel controls, self-tuning radios, pre-set buttons etc. The greatest source of internal distraction is passengers. In several cases, the passengers in question were children, who can be unpredictable in behaviour, lack appreciation of the stress a driver is under, and who may place high, emotive demands on the driver’s attention.

It was suggested in the previous chapter that internal distractions can be either controlled, or initiated, by the driver, or non-driver initiated. This is supported by the results from analysis of the database. Whilst a driver has some control over whether or not they engage in a mobile phone call, use their taxi radio, smoke, eat or drink etc., they do not have complete control over the behaviour of passengers, in particular children, or, for example, the movement of items in their vehicle, such as luggage. Distractions from sources such as these can be unpredictable and may coincide with driving situations that are particularly demanding.

In cases where mobile phones were identified as the source of distraction, the database does not indicate whether the call was driver initiated. A ringing phone can be difficult to ignore, particularly for those who drive for work, when the driver may perceive the call as important to their business. Similarly, when driving as part of their working life, drivers may feel the need to make

phone calls in order to increase their overall productivity. As discussed in the previous chapter, numerous studies have demonstrated a link between mobile phone use and increased crash risk.

Table A.1 – Sources of distraction in work-related accidents

External source	No.	Internal source (technology based)	No.	Internal source (non-technology based)	No.
Another vehicle	9	Mobile phones	7	Passengers	9
Road signs	3	Taxi/CB radio	3	Reaching for a dropped/falling item	5
Looking for buildings	3	Adjusting controls	2	Paper maps	3
Pedestrians	3	CD/stereo	1	Illness/pain	2
Lost/unfamiliar surroundings	3	Pager	1	Unknown	2
Unknown	3	Instrument panel	1	Smoking	2
Events in rear view mirror	2			Rubbing eyes	2
Off-road events	2			Securing a seatbelt	1
Cyclists	2			Door not shut properly	1
Events in right-hand road scene	1			Possible vehicle fault	1
Looking for turn-cues	1			Work-boots lodged between accelerator and brake pedal	1

Types of distraction

Within the work-related accident database, it can be difficult to determine the exact type(s) of distraction for each accident. The records do not specifically categorise accidents in this manner. It is sometimes possible to make judgements on the nature of the distraction. For example, where a driver is distracted by searching for road signs when lost it is reasonable to assume the distraction is visual and cognitive in nature, and unlikely to be biomechanical or auditory. However, in situations where a driver is distracted by passengers it is not clear whether this distraction is: visual, i.e. the driver turned to face their passenger and therefore was not observing the road; auditory, i.e. the

driver was startled or taken by surprise by a comment from their passenger; or cognitive, i.e. the conversation was mentally taxing and impinged on the driver's concentration.

It is unlikely that accident reports would be able to accurately determine the exact type of distraction. However, the types are useful for pre-crash research, where behaviour can be closely monitored to assess the different demands placed on drivers when performing secondary tasks. For example, researchers can evaluate the visual demand of a system using the occlusion technique (ISO, 2007), and therefore make a judgement on the potential visual distraction of a device, or assess cognitive distraction using workload questionnaires.

Distraction and the driving task

Considering the distraction sources of 'paper maps', 'road signs', 'looking for buildings' and 'lost/unfamiliar surroundings' (see Table A.1), as one category, route finding, it can be seen that this is the greatest overall source of distraction, accounting for 12 accidents. This provides further support for the research of Cnossen et al. (2004) (see above and Chapter 2). The further development of good interfaces for in-vehicle navigation systems may be fundamental in addressing this particular distraction source. However, if poorly designed, navigation system interfaces have the potential to exacerbate the problem. Particularly, as the results here demonstrate, since drivers seem willing to compromise driving performance for activities linked to route finding.

Distraction and accident type

Analysis of the work-related accident database shows that distraction related accidents differ in type to all accidents. 80% of distraction related accidents are either of the form rear end shunt or right of way violation (ROWV). Of these, rear end shunt is the greater category, at 44% (see figure A.7). In contrast, across all accidents, these types of accidents account for just 58%, with rear end shunts only 26% (see figure A.8).

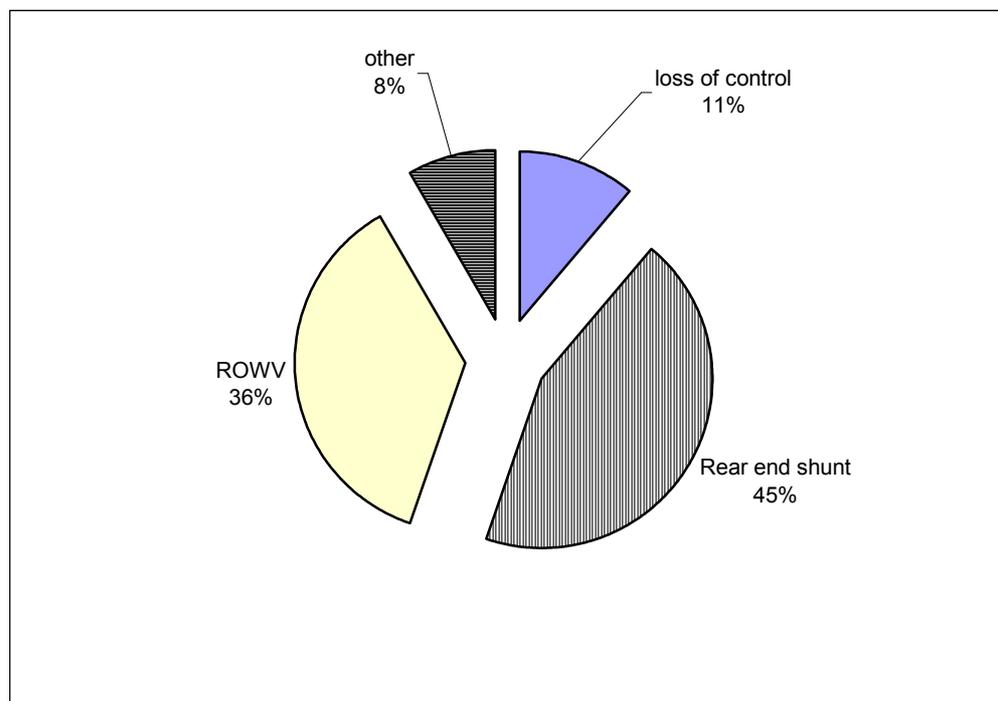


Figure A.7 Distraction related accident types

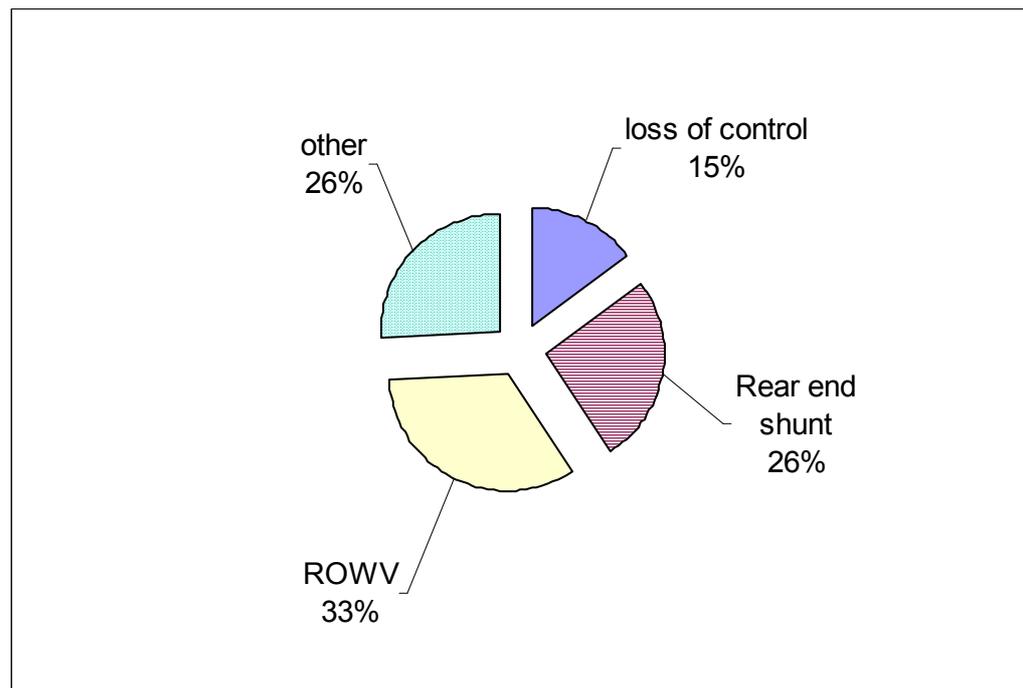


Figure A.8 All accident types

The differences between the types of accidents caused by distraction and all accidents provide researchers with some useful information. It is possible that, if distraction becomes a more significant cause of accidents, there will be an impact on the distribution of all accident types. This has implications for vehicle safety research, and may lead to increased development of safety systems that mitigate the effects of rear end shunt type collisions (Neale et al., 2005). The results are indicative of the reduced situational awareness of distracted drivers; drivers misjudge and are less aware of the road environment whilst engaged by distraction sources. Developments in driver support systems may seek to compensate for the impaired situational awareness of distracted drivers. For pre-crash analysis, the results provide support for the use of particular measures when assessing distraction. In simulator studies, measures of reaction times to a braking lead vehicle could be used with greater confidence, as a measure of distraction when engaging in a secondary task.

Summary

This appendix presented an analysis of distraction and work-related road traffic accidents that took place between 1996 and 2004 in the midlands region of the United Kingdom. All accidents in the database involved at least one vehicle driving on work-related business, including company car drivers, although fault, or blame, did not necessarily lie with the work-related driver. Accidents coded within the database as due, in some part, to distraction, accounted for 5.1% of all accidents. However, on closer inspection it was found that some of these accidents had been inaccurately categorised, and were more likely due to inattention rather than distraction. A revised figure suggested that 3.7% of accidents were caused, to some degree, by distraction. This figure is small in comparison with previous studies, and may be a result of the reasonably specific population under investigation. Key sources of distraction, such as passengers, may be underrepresented in the accidents contained within the database.

A greater number of internal distraction sources were coded compared with external distractions, with technology sources identified in 15% of distraction-related accidents. Of these, mobile phones accounted for seven accidents, the third greatest overall source of distraction. The analysis shows both internal and external distraction sources could be said to be driver-initiated and non-driver-initiated, contradicting previous suggestions (Lam, 2002). When considering four source categories (paper maps, road signs, looking for buildings, and lost/unfamiliar surroundings) under a single category of *route finding*, this becomes the greatest source of distraction. This emphasizes previous research, which suggested drivers may accept greater detriments in

driving performance in order to serve overall driving goals of reaching their destination (Cnossen et al., 2004).

An analysis of the types of accidents caused by distraction showed that the majority are of the rear end shunt or right of way violation type. This is further evidence of the reduced situational awareness of distracted drivers, and may be useful for designers of driver support systems. Such systems may pre-warn, or alert, a driver of a potential collision, or seek to minimise the impact of a collision.

Appendix B: Example Consent Forms

Study 1

INFORMED CONSENT FORM

I _____ (Full Name) _____ (Date of Birth) understand fully the procedures in this experiment, which have been explained to me.

I understand that as a participant I am guaranteed complete confidentiality and anonymity and although a record will be kept of my participation in the study, all data will be identified by number only. Therefore I consent to the written publication of results.

As parts of the trial will be recorded by video we would like to ask your permission to present, if necessary, some short video clips to our project partners.

Also the clips could be used for technical presentations and for teaching purposes. These video clips would only be used in a professional environment. Please tick the box below if you agree to your video being used by us:

I have been informed that there is no known expected discomfort or risk involved in my participation in this study (other than the task of driving), and have been asked about any medical conditions that might create a risk for me when I participate.

I have been informed that there are no “disguised” procedures in this study.

I have been informed of the purpose of the study and understand that after the experiment, the investigator will answer any questions regarding the procedures of this study.

I understand that whilst in control of the vehicle I am responsible for observing UK traffic regulations (and therefore speed limits). I accept liability for any traffic offences or violations, and take full responsibility for the consequences.

I am aware that the vehicle provided is comprehensively insured by Loughborough University.

I have been informed that I am free to withdraw from the study at any point and that I will be paid accordingly.

I confirm the following details are accurate:

- I have a valid licence for driving a manual car
- I have held this licence for at least 3 years.
- I do not have any other driving convictions or pending prosecutions
- I have not been drinking alcohol or taking drugs prior to driving.

I am aware that concerns about this study can be referred to the investigator, Mr Steven Bayer or his superior, Mr John Richardson.

I therefore give my consent to be a participant in this experiment.

Study 4

The purpose of this study is to evaluate how well the occlusion technique assesses the potential distraction effect of performing tasks on a navigation system when compared with data from simulator trials. The occlusion technique involves users performing tasks whilst their vision is intermittently obscured, in order to mimic the effect of glancing between a system and the road whilst driving.

This study comprises of two sessions, and you will be paid for your time at a rate of £8/hour. In order to take part in this study you must hold a valid driving licence for a manual car and have held this licence for at least two years.

As part of the evaluation you will be asked to use a driving simulator. During the drive you will be asked to carry out a number of in-car tasks with the controls. The experimenter will issue the instructions during the drive. You will be asked to follow a vehicle on the outside lane (right lane) during the drive on a dual-carriage way. The speed limit will be displayed on signs consistent with the Highway Code. You should drive at a distance you deem safe from the car you are following.

The experimenter will explain the equipment being used and make any necessary adjustments to ensure your comfort. You will be fully trained in the use of the navigation system as part of the study. There will be a series of tasks to ensure your familiarity with the various controls of the systems together with a practice drive to give you the opportunity to familiarise yourself with the driving simulator controls before the study commences.

Anecdotal reports suggest that some people may experience feelings of nausea when using driving simulators. If at anytime you feel in any form of discomfort, please inform the experimenter who will end the study immediately.

If you regularly suffer from/experience any of the following conditions it is advisable that you do not take part in the trial:

- **Migraine**
- **Blurred vision**
- **Motion sickness**

As a participant you will be guaranteed complete confidentiality and anonymity, with all recorded data from your participation identified by number only. The study will be recorded by video in order to monitor your interactions

with the controls of the car and the driving scene. The videos may be used at a later date to aid future analysis and may form some part of a public presentation.

This study is not a test of your driving abilities. If you have any questions please do not hesitate to ask. Your participation in this study is very much appreciated.

You may withdraw from this experiment at anytime.

I HAVE READ AND FULLY UNDERSTOOD THE ABOVE. MY PARTICIPATION IN THIS STUDY IS ENTIRELY VOLUNTARY.

Sign and Print name:

Date:

Appendix C: Participant questionnaires

Study 1

1. Personal details

1.1. First Name

1.2. Surname

1.3. Gender

Male

Female

1.4. What title would you prefer us to use? (Mr / Mrs / Ms / Other
.....)

1.5. Date of Birth

1.6. What is your home address

.....
.....
.....
.....

1.7. Email

1.8. Daytime Phone No.

2. Driving experience details

2.1. What year did you pass your driving test?

.....

2.2. Approximately how many miles do you drive per year?

0 – 1,000 miles

1,001 – 5,000 miles

5,001 – 10,000 miles

10,001 – 15,000 miles

15,001 – 20,000 miles

20,001 + miles

2.3. How often do you drive on the following roads?

Type of road	Daily	Weekly	Monthly	Never
Rural & urban estate roads				
'A' roads				
'B' roads				
Dual-carriageways				
Motorways				

3.4. What type of vehicle do you normally drive?

Make (e.g. Ford)

.....

Model (e.g. Fiesta)

.....

Year of registration (e.g. X
Reg).....

3.5. Has your car got a CD player?

Yes No

3.6. Has your car got a navigation system?

Yes No

3.7. Has your car got a menu-based computer system?

Yes No

Study 4

1. What age group do you belong to?

18 – 25 26 – 35 36 – 45 46 – 55 55 +

2. Gender? MALE FEMALE

3a). Approximately how many miles (to the nearest 1000) have you driven in the past 12 months?

.....

3b). Would you say this mileage is typical? YES NO

4. How many years have you held your full driving licence?.....years

5. On average, how many days per week do you currently drive?

Please tick appropriate box

<input type="checkbox"/>	1
<input type="checkbox"/>	2
<input type="checkbox"/>	3
<input type="checkbox"/>	4
<input type="checkbox"/>	5
<input type="checkbox"/>	6
<input type="checkbox"/>	7

6 a) Have you ever used an in-car navigation system? YES NO

b) If yes, what was the make of the navigation system?

<input type="checkbox"/>	Manufacturer known:
<input type="checkbox"/>	Manufacturer unknown, supplied with vehicle:
<input type="checkbox"/>	Manufacturer unknown, not supplied with vehicle:

7. Are you right or left-handed? RIGHT LEFT

Appendix D: Raw data

Study 1

TTTstatic - Navigation task – System A						
Participant	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Average
2b	13.76	13.24	9.19	6.97	12.70	11.17
3b	19.17	15.64	18.09	8.67	24.56	17.23
4b	11.13	8.62	9.86	8.22	7.91	9.15
5b	6.99	7.67	23.82	6.68	7.35	10.50
6b	11.63	10.56	5.91	7.11	9.01	8.84
7b	13.51	13.13	24.23	14.41	13.93	15.84
8b	7.14	7.80	5.54	6.28	9.13	7.18
1c	12.84	11.06	12.75	11.22	10.99	11.77
2c	7.80	5.95	5.00	5.51	6.88	6.23
3c	11.09	6.49	8.80	8.19	6.87	8.29
4c	10.64	8.56	7.79	3.34	9.02	7.87
5c	14.75	16.33	19.84	14.97	25.52	18.28
6c	6.55	1.43	2.27	9.32	10.07	5.93
7c	19.45	19.30	18.23	17.55	16.81	18.27
8c	10.04	7.52	7.32	6.49	8.33	7.94

TTTstatic - Navigation task – System B						
Participant	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Average
1a	16.18	11.19	12.73	7.39	9.57	11.41
2a	14.50	9.08	15.55	14.92	12.69	13.35
4a	12.73	10.50	8.74	8.49	11.30	10.35
5a	15.01	8.48	8.97	6.68	7.13	9.25
6a	13.05	9.02	10.31	13.03	8.43	10.77
8a	8.03	7.10	7.83	12.72	5.72	8.28
1d	10.48	10.55	12.67	10.29	9.02	10.60
2d	8.88	7.27	8.07	5.18	6.08	7.10
3d	9.38	8.69	8.69	15.19	6.91	9.77
4d	15.27	10.97	14.31	9.87	11.73	12.43
5d	45.12	11.97	15.83	15.43	10.54	19.78
6d	15.99	11.14	7.81	8.87	9.85	10.73
7d	14.81	13.03	19.56	17.17	9.69	14.85
8d	10.60	8.88	8.51	10.02	16.12	10.83

TTTstatic - Sound task – System A						
Participant	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Average
2b	16.20	19.30	9.73	13.51	11.17	13.98
3b	28.45	18.64	20.19	6.12	12.73	17.23
4b	27.78	22.97	15.66	13.30	8.18	17.58
5b	10.46	10.59	8.27	15.14	7.21	10.33
6b	11.03	7.99	14.20	10.16	14.53	11.58
7b	9.76	9.73	7.87	7.90	8.23	8.70
8b	9.46	9.75	7.82	10.57	10.05	9.53
1c	10.50	15.01	10.84	12.97	15.37	12.94
2c	12.20	7.59	13.40	10.43	6.88	10.10
3c	13.20	10.13	11.62	10.73	13.33	11.80
4c	15.43	12.22	42.87	13.99	17.08	20.32
5c	27.27	16.90	19.79	14.65	17.85	19.29
6c	12.73	11.58	10.67	8.36	8.37	10.34
7c	19.35	15.77	11.03	18.65	13.54	15.67
8c	8.37	9.46	7.82	10.14	8.50	8.86

TTTstatic - Sound task – System B						
Participant	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Average
1a	29.43	23.88		15.29	28.37	24.24
2a	32.27	21.49	35.49	35.41	33.45	31.62
4a	21.43	21.81	34.48	16.25	46.85	28.16
5a	26.28	36.85	27.55	29.19	21.86	28.35
6a	27.15	16.43	28.50	20.77	17.19	22.01
8a	15.59	15.68	21.19	19.83	13.56	17.17
1d	30.78	37.79	26.28	24.73	25.29	28.97
2d	18.17	16.23	22.07	21.89	15.75	18.82
3d	23.32	19.47	20.57	18.77	16.41	19.71
4d	23.58	37.43	26.40	25.44	23.62	27.29
5d	56.54	21.81	48.53	24.97	30.05	36.38
6d	33.43	26.87	30.17	20.31	29.06	27.97
7d	24.40	36.59	46.19	38.01	36.40	36.32
8d	20.05	26.13	25.45	21.47	21.83	22.99

TSOT - navigation task – System A						
Participant	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Average
2b	14.28	13.26	15.48	10.11	9.06	12.44
5b	8.39	6.04	11.67	7.77	5.95	7.96
6b	12.4	10.5		10.33	9.46	10.67
1c	5.33	6.03	4.75	3.41	4.57	4.82
2c		7.88	8.19	5.52	7.54	7.28
3c	11.26	4.53	6.57	11.61	5.12	7.82
4c		9.65		10.36	9.83	9.95
5c	18.16	11.2	11.39	8.88	16.72	13.27
6c			7.45		7.56	7.51
7c	12.08	8.55	5.88	2.97	17.15	9.33
8c	10.18	9.02	10.05	7.03		9.07

TSOT - navigation task – System B						
Participant	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Average
2a		11.65	16.26	11.65	7.24	11.70
5a	11.43	7.29	12.89	10.05	3.90	9.11
6a	14.35	8.59	7.06	10.62	5.61	9.25
1d	9.85	5.78	11.57	11.00	5.14	8.67
2d	9.72	5.65	8.49	9.04	5.93	7.77
3d	14.57	15.72	10.06	9.68	4.15	10.84
4d		9.08	6.85	9.50	3.41	7.21
5d	9.73	9.96	8.59	11.97	4.78	9.00
6d	11.32	8.53	10.43	19.06	4.66	10.80
7d	15.81	7.55	8.73	23.51	9.10	12.94
8d	5.45	7.48	12.31	10.07	7.82	8.63

TSOT - sound settings task – System A						
Participant	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Average
2b	8.31	10.92	4.79	9.75	10.27	8.81
5b	7.35	7.14	5.13	6.99	6.27	6.58
6b	9.68	8.54	8.67	6.83	6.77	8.10
1c	6.84	7.36	7.08	8.37	8.74	7.68
3c	5.80	5.58	4.96	5.36	8.22	5.98
4c	7.48	17.43	17.86	10.13	13.83	13.35
5c	10.29	9.48	7.84	12.06	6.70	9.27
6c	9.58	2.73	9.14	8.93	9.24	7.92
7c	17.58	13.14	14.31	10.59	15.28	14.18
8c	4.75	3.95	5.15	4.23	3.92	4.40

TSOT - Sound settings task – System B						
Participant	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Average
2a	22.17	17.10	26.52	17.58	19.34	20.54
5a	11.52	13.55	0.00	10.12	11.73	9.38
6a	20.71	11.62	13.38	12.78	8.41	13.38
1d	17.59	14.35	12.96	20.24	24.82	17.99
3d	11.04	8.40	8.07	10.60	9.40	9.50
4d	24.15	9.89	10.42	20.04	11.55	15.21
5d		11.11		10.25		10.68
6d	15.81	12.48	17.82	12.99	10.95	14.01
7d	16.43	11.47	13.92	15.10	14.84	14.35
8d	11.29	12.90	18.22	12.67	12.23	13.46

On road results - Navigation Task – Control – System A					
Participant	Total Task Time	Total Glance Time	Prop. Eyes off road	Frequency	Single glance time
2b	11.24	8.52	75.80	8.00	1.07
3b	43.80	27.64	63.11	21.00	1.32
5b	19.56	11.64	59.51	12.00	0.97
6b	25.16	18.96	75.36	15.00	1.26
7b	40.96	12.00	29.30	13.00	0.92
8b	15.72	7.96	50.64	11.00	0.72
1c	13.32	8.32	62.46	7.00	1.19
2c	12.24	9.28	75.82	9.00	1.03
3c	13.84	8.60	62.14	9.00	0.96
4c	20.88	12.52	59.96	13.00	0.96
5c	26.00	19.08	73.38	15.00	1.27
6c	18.56	7.72	41.59	8.00	0.97
8c	17.28	12.80	74.07	9.00	1.42

On road results - Navigation Task – System B					
Participant	Total Task Time	Total Glance Time	Prop. Eyes off road	Frequency	Single glance time
1a	26.72	13.44	50.30	12.00	1.12
2a	20.88	15.84	75.86	13.00	1.22
4a	31.48	14.92	47.40	13.00	1.15
5a	25.88	12.52	48.38	13.00	0.96
6a	16.92	13.32	78.72	10.00	1.33
7a	19.16	11.58	60.44	12.00	0.97
8a	24.20	18.96	78.35	14.00	1.35
1d	16.96	11.12	65.57	8.00	1.39
2d	15.00	10.44	69.60	8.00	1.31
3d	12.16	10.00	82.24	7.00	1.43
5d	35.92	24.52	68.26	16.00	1.53
6d	35.56	14.92	41.96	17.00	0.88
7d	23.20	7.20	31.03	10.00	0.72
8d	3.88	2.00	51.55	2.00	1.00

On road results - Sound Task – System A					
Participant	Total Task Time	Total Glance Time	prop eyes off road	Frequency	Single glance time
2b	14.36	10.12	70.47	10.00	1.01
3b	20.44	16.24	79.45	10.00	1.62
4b	18.80	10.92	58.09	9.00	1.21
5b	18.08	7.44	41.15	8.00	0.93
7b	10.84	6.00	55.35	6.00	1.00
8b	11.44	8.92	77.97	9.00	0.99
1c	15.04	9.12	60.64	7.00	1.30
2c	10.76	7.24	67.29	6.00	1.21
3c	9.04	7.24	80.09	7.00	1.03
4c	22.20	8.12	36.58	11.00	0.74
5c	34.72	22.60	65.09	18.00	1.26
6c	14.88	5.72	38.44	6.00	0.95
7c	16.84	9.84	58.43	9.00	1.09
8c	11.76	7.24	61.56	7.00	1.03

On road results - Sound Task – System B					
Participant	Total Task Time	Total Glance Time	Prop. eyes off road	Frequency	Single glance time
1a	28.20	11.84	41.99	11.00	1.08
3a	55.20	27.36	49.57	22.00	1.24
4a	37.24	18.68	50.16	18.00	1.04
5a	32.60	11.04	33.87	13.00	0.85
6a	35.96	11.88	33.04	11.00	1.08
7a	29.92	8.32	27.81	10.00	0.83
8a	32.52	14.56	44.77	16.00	0.91
1d	42.52	20.68	48.64	13.00	1.59
2d	27.40	19.36	70.66	13.00	1.49
3d	21.36	10.24	47.94	9.00	1.14
5d	29.48	16.80	56.99	14.00	1.20
6d	26.40	10.52	39.85	13.00	0.81
7d	6.56	4.20	64.02	6.00	0.70
8d	20.16	15.52	76.98	9.00	1.72

Study 2

Task	Observed task times for system 1									Mean	Median	Standard deviation	
Switch on radio	2.54	2.81	3.51	1.68	1.83	1.65	1.98	1.67	2.22	1.87	2.18	1.93	0.61
Tune into radio 1	3.59	3.53	3.54	2.87	3.27	3.48	3.19	3.21	3.98	3.56	3.42	3.51	0.30
Save frequency to preset 1	3.23	4.44	3.98	3.40	2.51	2.43	3.10	2.01	2.87	3.09	3.11	3.10	0.72
Increase volume	2.54	2.45	2.13	2.54	2.18	2.06	1.87	2.34	2.56	2.54	2.32	2.40	0.25
Tune into BBC Notts	4.99	5.43	4.87	4.32	4.63	4.54	5.45	3.65	4.76	5.21	4.79	4.82	0.55
Save to preset 2	3.66	2.85	3.90	3.02	3.01	3.79	2.48	3.60	3.67	3.87	3.39	3.63	0.50
Decrease volume	2.54	1.99	3.01	2.10	1.64	2.23	1.86	2.43	2.54	2.99	2.33	2.33	0.46
Insert tape	1.88	1.82	1.99	0.97	1.15	1.65	1.21	2.21	1.43	1.56	1.59	1.61	0.40
Increase bass to full	3.71	3.98	4.01	4.32	3.50	3.51	3.08	3.21	3.31	3.21	3.58	3.51	0.41
Eject tape	2.35	2.66	2.01	2.21	1.91	1.62	1.85	1.89	1.87	1.98	2.04	1.95	0.30
Switch off radio	2.03	2.43	1.87	2.01	1.98	1.37	2.31	2.22	1.87	1.99	2.01	2.00	0.29

Task	Observed task times for system 1									Mean	Median	Standard deviation	
Switch on radio	2.01	4.45	4.22	4.09	3.99	4.32	1.98	2.01	2.45	2.98	3.25	3.49	1.06
Tune into radio 1	9.76	6.16	6.78	8.51	5.67	6.32	8.43	9.76	8.50	8.45	7.83	8.44	1.49
Save frequency to preset 1	5.13	4.21	4.76	3.03	4.78	4.06	4.90	5.13	5.01	3.90	4.49	4.77	0.68
Increase volume	2.52	2.00	3.81	3.00	2.87	2.67	2.98	2.52	2.90	2.76	2.80	2.82	0.46
Tune into BBC Notts	11.41	9.68	10.45	9.64	8.91	9.41	9.45	10.41	9.12	8.54	9.70	9.55	0.85
Save to preset 2	3.53	4.10	3.56	3.51	3.99	3.98	3.43	3.53	3.54	3.78	3.70	3.55	0.25
Decrease volume	2.75	4.34	4.01	1.82	4.56	4.10	2.67	2.75	3.21	2.99	3.32	3.10	0.89
Insert tape	3.71	2.02	3.51	1.87	2.54	1.98	3.90	3.71	3.11	3.40	2.98	3.26	0.80
Increase bass to full	3.70	2.43	4.50	4.15	2.87	2.87	3.43	3.70	3.23	4.10	3.50	3.57	0.66
Eject tape	2.42	1.93	2.67	2.54	2.32	2.30	2.78	2.42	2.98	2.21	2.46	2.42	0.30
Switch off radio	2.31	2.89	2.54	2.65	2.11	2.46	2.34	3.14	2.87	2.25	2.56	2.50	0.33

Study 4**System A task 1 - static trials**

<i>Participant</i>	<i>Trial 1</i>	<i>Trial 2</i>	<i>Trial 3</i>	<i>Trial 4</i>	<i>Trial 5</i>	<i>Average</i>
1	57.78	41.71	53.19	67.45	50.4	54.106
2	43.6	43.33	56.79	38.47	41.13	44.664
3	52.56		75.24	64.62	62.59	63.7525
4	57.87	53.59	39.46	44.77	52.6	49.658
5	48.64		42.52	55.66	46.26	48.27
6	67.77	49.18	72.22	68.44	57.82	63.086
7	52.69	38.88	54.36	45.31	43.65	46.978
8	74.79	52.29	60.21	50.67	56.47	58.886
9	53.56	54.4	57.46	60.39		56.4525
10	44.37	44.82	50.49	46.39	50.71	47.356
11	63.08	44.86	58.77	63.99	55.75	57.29
12	46.35	43.02	44.64	34.92	43.47	42.48

System A task 1 - occluded trials

<i>Participant</i>	<i>Trial 1</i>	<i>Trial 2</i>	<i>Trial 3</i>	<i>Trial 4</i>	<i>Trial 5</i>	<i>Average</i>	<i>R</i>
1	28.319	31.521	25.337	39.026	55.477	35.936	0.664
2	31.523	31.523	27.019		35.683	31.437	0.704
3		36.027	30.021	44.226	55.796	41.5175	0.651
4	46.532	37.663	24.016	36.03		36.06025	0.726
5	38.144		27.018	36.025	36.028	34.30375	0.711
6	37.528	29.166	31.521			32.73833	0.519
7	39.031	27.017	30.002	33.029	39.032	33.6222	0.716
8	33.04	31.529	27.017			30.52867	0.518
9	49.532	43.534		73.555	66.043	58.166	1.030
10		31.521	33.028	39.03	42.033	36.403	0.769
11	42.699	67.545		67.545		59.263	1.034
12	37.525	37.527	30.028	30.028	39.03	34.8276	0.820

System A task 2 - static trials

<i>Participant</i>	<i>Trial 1</i>	<i>Trial 2</i>	<i>Trial 3</i>	<i>Trial 4</i>	<i>Trial 5</i>	<i>Average</i>
1	26.68	13.41	18.49		17.01	18.8975
2	20.47	11.79	18.36	13.95	19.93	16.9
3	20.02	12.96	34.92	16.47	15.43	19.96
4	20.38	18	23.4	19.03	16.33	19.428
5	16.51	12.91		17.64	12.46	14.88
6	18.49	12.69		14.49	14.8	15.1175
7	21.64	14.35	16.51	14.08	16.11	16.538
8	19.21	14.98		12.15	13.68	15.005
9	23.04	8.68	17.41	9.99	15.12	14.848
10	18.9	12.64	15.61		13.05	15.05
11		13.09	16.29		18.63	16.0033
12	19.62	12.64	15.48	9.09	9.22	13.21

System A task 2 - occluded trials

<i>Participant</i>	<i>Trial 1</i>	<i>Trial 2</i>	<i>Trial 3</i>	<i>Trial 4</i>	<i>Trial 5</i>	<i>Average</i>	<i>R</i>
1	16.401	15.01	16.512		11.778	14.92525	0.790
2	12.008	8.981	6.004	10.45	9.006	9.2898	0.550
3	12.008	15.016	7.508	10.507	10.511	11.11	0.557
4	15.004	13.509	9.014	12.016	12.008	12.3102	0.634
5	15.015	10.51	7.511	12.008	9.01	10.8108	0.727
6	12.015	11.93	8.28		9.006	10.30775	0.682
7	13.512	13.517	10.791	10.507	12.014	12.0682	0.730
8	15.016	12.013	7.51	10.507	15.012	12.0116	0.801
9	12.009	10.507	10.507	12.657	13.509	11.8378	0.797
10	12.013	13.516	9.006	12.009	10.514	11.4116	0.758
11	19.514	18.02	12.011	22.519	13.509	17.1146	1.069
12	10.509	9.005	7.508	12.011	9.009	9.6084	0.727

System A task 3 - static trials

<i>Participant</i>	<i>Trial 1</i>	<i>Trial 2</i>	<i>Trial 3</i>	<i>Trial 4</i>	<i>Trial 5</i>	<i>Average</i>
1	10.35		8.37	7.74	6.97	8.3575
2	8.01	10.62	6.97	7.42	10.08	8.62
3	7.33	6.97	8.95	7.47		7.68
4	6.3	10.21	7.38	6.7	6.79	7.476
5	6.66	6.07	5.58	6.3	4.14	5.75
6	8.37	10.98	8.5	10.08	8.14	9.214
7	10.75	8.95	9.49	13.23	9.04	10.292
8	7.11	12.91	6.75	8.68	5.4	8.17
9	9.4	18.94	6.48	8.77	7.96	10.31
10	24.21	9.72	10.71	9.31	12.64	13.318
11	8.05	11.25	9.04	7.56	7.24	8.628
12	7.24	9.63	5.98	6.52	5.8	7.034

System A task 3 - occluded trials

<i>Participant</i>	<i>Trial 1</i>	<i>Trial 2</i>	<i>Trial 3</i>	<i>Trial 4</i>	<i>Trial 5</i>	<i>Average</i>	<i>R</i>
1	7.505	9.011	8.919	4.503	6.007	7.189	0.860
2	6.008	9.01	4.507	4.508	6.005	6.0076	0.697
3	7.505	7.51	4.81	3.002	8.85	6.3354	0.825
4	9.297	12.008	7.512	6.004	6.003	8.1648	1.092
5	9.007	6.006	6.735	4.503	5.131	6.2764	1.092
6		10.506		4.503	7.504	7.504333	0.814
7	10.513	9.011	7.28	4.503		7.82675	0.760
8	6.007	7.505	5.727	4.503	6.009	5.9502	0.728
9	13.517	9.005	10.507	6.011	12.01	10.21	0.990
10	13.468	9.008	7.505	6.004	7.507	8.6984	0.653
11	15.016	10.506	10.342	11.898	7.345	11.0214	1.277
12	9.706	6.004	7.51	7.51	4.511	7.0482	1.002

System B task 1 - static trials

<i>Participant</i>	<i>Trial 1</i>	<i>Trial 2</i>	<i>Trial 3</i>	<i>Trial 4</i>	<i>Trial 5</i>	<i>Average</i>
1	22.63	21.24	24.7	23.08	25.69	23.468
2	19.8	13.18	17.32	14.4		16.175
3	19.3	19.57	25.6	23.76	25.2	22.686
4		14.13	12.55	18.09	19.62	16.0975
5	17.1	13.32	18.72	15.43	9.85	14.884
6		16.38	17.37	18.31		17.3533
7	25.24	28.17	22	22.18		24.3975
8		21.46	18.94	18.22	26.41	21.2575
9	22.5	20.75	15.75	18.09	26.55	20.728
10	27.31	22.86	23.4	29.83	33.84	27.448
11	31.14	32.26	26.73	25.42	30.78	29.266
12	11.97	13.95	13.5	12.42	13.86	13.14

System B task 1 - occluded trials

<i>Participant</i>	<i>Trial 1</i>	<i>Trial 2</i>	<i>Trial 3</i>	<i>Trial 4</i>	<i>Trial 5</i>	<i>Average</i>	<i>R</i>
1	21.014	18.011	18.02	22.519	25.519	21.017	0.896
2	16.511	21.018	10.507	15.01	19.516	16.512	1.021
3	22.382	25.536	16.511	18.02	27.022	21.894	0.965
4	21.019	16.518	13.509	16.519		16.891	1.049
5	10.373	6.004	13.441	15.01	10.515	11.069	0.744
6		13.828	16.511	15.014	16.512	15.466	0.891
7		16.515	19.515	21.013	24.017	20.265	0.831
8	18.021	13.509	15.01	24.022		17.641	0.830
9	30.022	29.793	16.511		19.513	23.960	1.156
10	15.01		17.845	36.263		23.039	0.839
11	45.018		27.018			36.018	1.231
12	25.518	18.013	13.467	12.008	16.516	17.104	1.302

System B task 2 - static trials

<i>Participant</i>	<i>Trial 1</i>	<i>Trial 2</i>	<i>Trial 3</i>	<i>Trial 4</i>	<i>Trial 5</i>	<i>Average</i>
1		11.97	15.43	13.45	6.21	11.765
2	7.11	7.11	8.5	13.18	6.75	8.53
3		8.05		10.26	4	7.43667
4	8.37	9.85	13.95	10.84	5.4	9.682
5	6.12	5.44	13.95	9.04	4.27	7.764
6	6.79	5.49	11.38	9.76		8.355
7	13.45	15.39	13.18	9.04	10.66	12.344
8	8.77	6.52	10.8	11.61	5.94	8.728
9	6.97	8.05	18.76	9.9	4.72	9.68
10	9.85	9.99	13.45	12.64	7.6	10.706
11	9.9	7.78	13.05	14.46	8.86	10.81
12	7.24	5.35	8.82	11.79	5.76	7.792

System B task 2 - occluded trials

<i>Participant</i>	<i>Trial 1</i>	<i>Trial 2</i>	<i>Trial 3</i>	<i>Trial 4</i>	<i>Trial 5</i>	<i>Average</i>	<i>R</i>
1	4.502	7.512		3.002		5.005	0.425
2	6.004	12.016	6.01		3.002	6.758	0.792
3		12.015	4.505	6.008	7.507	7.509	1.010
4	4.503	12.01	4.504		9.01	7.507	0.775
5	4.509	8.971	2.882	4.253	6.004	5.324	0.686
6	4.51	7.397	4.503	4.511	7.507	5.686	0.681
7	9.013		10.512	7.481	6.004	8.253	0.669
8	4.509	9.01	4.506	5.941	7.504	6.294	0.721
9	9.009	12.009	6.007	6.003	7.508	8.107	0.838
10	4.503	13.508	12.011	9.903	9.006	9.786	0.914
11	6.005	14.273	4.503	9.006	9.012	8.560	0.792
12	6.009	9.008	4.507	6.006	6.01	6.308	0.810

System B task 3 - static trials

<i>Participant</i>	<i>Trial 1</i>	<i>Trial 2</i>	<i>Trial 3</i>	<i>Trial 4</i>	<i>Trial 5</i>	<i>Average</i>
1	6.25	4.05	4.27	2.88	3.6	4.21
2	2.97	3.33	3.46	5.26	6.12	4.228
3	1.89		2.29	2.2	1.57	1.9875
4	3.15	3.37	2.88	2.79	3.24	3.086
5	1.98	1.8	1.62		1.66	1.765
6	2.38	2.56	2.34	3.33	2.34	2.59
7	6.12	4.9	4.54	4.5	3.87	4.786
8	2.92	3.15	3.42	2.92	2.56	2.994
9	3.6	6.39	4.32	3.73	3.91	4.39
10	4.27		4.23	3.64	4.9	4.26
11	3.01	2.25	2.56	2.7	2.29	2.562
12	4.41	1.57	1.8	1.8	1.57	2.23

System B task 3 - occluded trials

<i>Participant</i>	<i>Trial 1</i>	<i>Trial 2</i>	<i>Trial 3</i>	<i>Trial 4</i>	<i>Trial 5</i>	<i>Average</i>	<i>R</i>
1	2.325	1.501	1.635		1.504	1.741	0.414
2	3.009	3.007	2.987	1.501	4.636	3.028	0.716
3	1.504	1.501	1.508	1.505	1.509	1.505	0.757
4	1.501	1.507		1.751	1.79	1.637	0.531
5	1.5		1.508	1.506		1.505	0.853
6	1.5	1.5		1.506	1.503	1.502	0.580
7	5.449	3.002	3.009	3.002	3.001	3.493	0.730
8	1.5	3.01	4.505	1.505	1.504	2.405	0.803
9	1.508	1.5	1.504	2.193	0.88	1.517	0.346
10	4.51	6.004	1.75	3.002	3.005	3.654	0.858
11	2.937	1.5	1.503	1.508	1.502	1.790	0.699
12	6.007	1.502	5.965	1.506	2.748	3.546	1.590